

BIO2450L

Genetics Laboratory

Manual



Image credit: CC BY-SA 3.0 en:User:Cburnett (author)

Prof. Christopher Blair
Department of Biological Sciences
New York City College of Technology
cblair@citytech.cuny.edu

Labs also available online: <https://openlab.citytech.cuny.edu/geneticslab/>

Updated 8/4/21



Prof. Christopher Blair
cblair@citytech.cuny.edu

Laboratory Schedule

Handouts will be used for all laboratories in this course. These will be posted to OpenLab site (<https://openlab.citytech.cuny.edu/geneticslab/>) and must be printed out and brought to class.

Week 1	Introduction to Genetics Probability, Mendelian Genetics, chi-square and measurements, controls	Date
Week 2	Cytogenetics and Karyotyping Chromosomal alterations and human disease	Date
Week 3	Mitosis & Meiosis <i>Sordaria</i> recombination and genetic crosses <i>Homework 1: Cytogenetics write-up due</i>	Date
Week 4	Quiz 1 – Introduction, Cytogenetics, Mitosis, Meiosis	
	Monohybrid and Dihybrid Crosses Fruit fly (<i>Drosophila</i>) genetics Lab 1 Population Genetics Human blood type frequencies Lab 1 <i>Homework 2: Sordaria write-up due</i>	Date
Week 5	Monohybrid and Dihybrid Crosses Fruit fly (<i>Drosophila</i>) genetics Lab 2 Population Genetics Human blood type frequencies Lab 2	Date
Week 6	Monohybrid and Dihybrid Crosses Fruit fly (<i>Drosophila</i>) genetics Lab 3 Simulating Population Genetic Processes Genetic drift, mutation, gene flow, natural selection <i>Homework 3: Blood typing and population genetics write-up due</i>	Date
Week 7	Monohybrid and Dihybrid Crosses Fruit fly (<i>Drosophila</i>) genetics Lab 4 Lab Review (Crosses and Population Genetics) Introduction to pipetting	Date
	Quiz 2 – Inheritance and Population Genetics	



	<p align="center">DNA profiling in forensic science Pipetting and DNA fingerprinting Lab 1 (restriction digests)</p> <p align="center"><i>Homework 4: Genetic crosses write-up due</i></p>	Date
Week 9	<p align="center">DNA profiling in forensic science DNA fingerprinting Lab 2 (gel electrophoresis and analysis)</p>	Date
Week 10	Quiz 3 – DNA Fingerprinting	
	<p align="center">Molecular Genetics, PCR, and Genotyping PTC genetics and GMO Lab 1 (DNA extraction and PCR)</p> <p align="center"><i>Homework 5: DNA fingerprinting write-up due</i></p>	Date
Week 11	<p align="center">Molecular Genetics, PCR, and Genotyping PTC genetics and GMO Lab 2 (restriction enzyme digestion and gel electrophoresis)</p>	Date
Week 12	<p align="center">Molecular Genetics, PCR, and Genotyping PTC genetics and GMO Lab 3 (lecture, data analysis, and review questions)</p>	Date
Week 13	<p align="center">DNA Barcoding and Evolutionary Genetics DNA extraction and PCR from fish samples</p>	Date
Week 14	<p align="center">DNA Barcoding and Evolutionary Genetics Gel electrophoresis, BLAST, multiple sequence alignment and phylogenetic inference</p>	Date
Week 15	Quiz 4 – Molecular and Evolutionary Genetics	
	Full Laboratory Report Due (PTC, GMO <u>OR</u> Barcoding)	



Table of Contents

Lab 1: Probability, Statistics, and Measurements

Background.....	5
Probability and statistics.....	6
Measurements and dilutions.....	12
Biological controls.....	13

Lab 2: Cytogenetics and Karyotyping

Background.....	15
Activity.....	20
Cytogenetics Report.....	22
Preparation for Lab 3.....	23

Lab 3: Meiosis and Genetic Recombination

Background.....	25
Activity.....	30
Review questions.....	33

Labs 4-7: Patterns of Inheritance: Fruit Flies, Blood Types, Population Genetics and Hardy-Weinberg

Lab overview.....	35
Fly activity for Week 4.....	36
Fly activity for Week 5.....	37
Fly activity for Week 6.....	38
Fly activity for Week 7.....	39
Background: human blood types and Hardy-Weinberg.....	43
Blood type activity (Weeks 4-5).....	50
Review questions.....	55
Simulating population genetic processes (Week 6).....	58

Labs 8-9: DNA Fingerprinting

Background.....	74
Activity (Week 8).....	77
Activity (Week 9).....	82
Review questions.....	86

Labs 10-12: Molecular Genetics, PCR, and Genotyping

Overview.....	89
PTC background.....	92
PTC activity (Week 10).....	94
PTC activity (Week 11).....	96
PTC activity (Week 12).....	97
PTC review questions.....	98
GMO background.....	101
GMO activity (Week 10).....	102
GMO activity (Week 11).....	104
GMO activity (Week 12).....	105
GMO review questions.....	105

Labs 13-14: Introduction to DNA Barcoding and Evolutionary Genetics

Background.....	108
Activity (Week 13).....	109
Activity (Week 14).....	110
Review questions.....	116



Probability, Statistics, Measurements and Controls

The goal of this laboratory is to provide an introduction to basic concepts and techniques commonly used to study genetics. The first portion of the lab is devoted to the concept of probability and how we can use basic probability theory to understand genetic concepts and **crosses**. Genetic crosses are commonly used to study patterns of inheritance of **traits**. A trait, or a character, is generally any observable phenotypic characteristic of an organism such as eye color, skin color, height, etc. Gregor Mendel, often considered the father of genetics, performed many genetic crosses to quantify patterns of inheritance in pea plants. Based on the results of his experiments he came up with three laws:

Law 1: Law of Segregation

Alleles in diploid individuals separate during the process of gamete formation (meiosis).

Remember that a diploid cell contains two sets of chromosomes, one from the father and one from the mother. Thus, each gene will contain two alleles. The alleles can either be the same (homozygous) or different (heterozygous). For example, if we assume that pea pod color (green versus yellow) is controlled by a single gene with two alleles (R and r), RR and rr would represent homozygotes and Rr would indicate a heterozygote. During gamete formation, only one of the two alleles will be passed on to the sperm or egg. In other words, the two alleles *segregate* from one another (see Fig. 1).

Law 2: Law of Independent Assortment

Different genes randomly sort their alleles during the process of gamete formation (meiosis).

For example, going back to Mendel's experiments with pea plants, suppose we are working with two genes we will call Gene 1 and Gene 2. Gene 1 controls pea pod color and consists of two alleles ($R = \text{green}$, $r = \text{yellow}$). We assume that the R allele is **dominant**, meaning that RR and Rr genotypes produce green pods and rr genotypes produce yellow pods. Now assume that Gene 2 controls seed pod shape and also contains two alleles ($Y = \text{constricted}$, $y = \text{round}$). Assume that Y is dominant over y , such that YY and Yy genotypes produce constricted pods and yy genotypes produce round pods. Mendel's Law of Independent Assortment states that the alleles at these different genes will sort independently of one another during gamete formation. In other words, the R allele will not always be associated with the Y allele and the r allele will not always be associated with the y allele in each sperm or egg cell. All combinations of alleles are possible (Fig. 1).

Law 3: Law of Dominance

A heterozygous individual will express the phenotypic characteristics of the dominant allele.

For example, in our green versus yellow plant example, we say that the green allele (*R*) is dominant to the yellow allele (*r*) because both *RR* and *Rr* plants demonstrate the green phenotype.

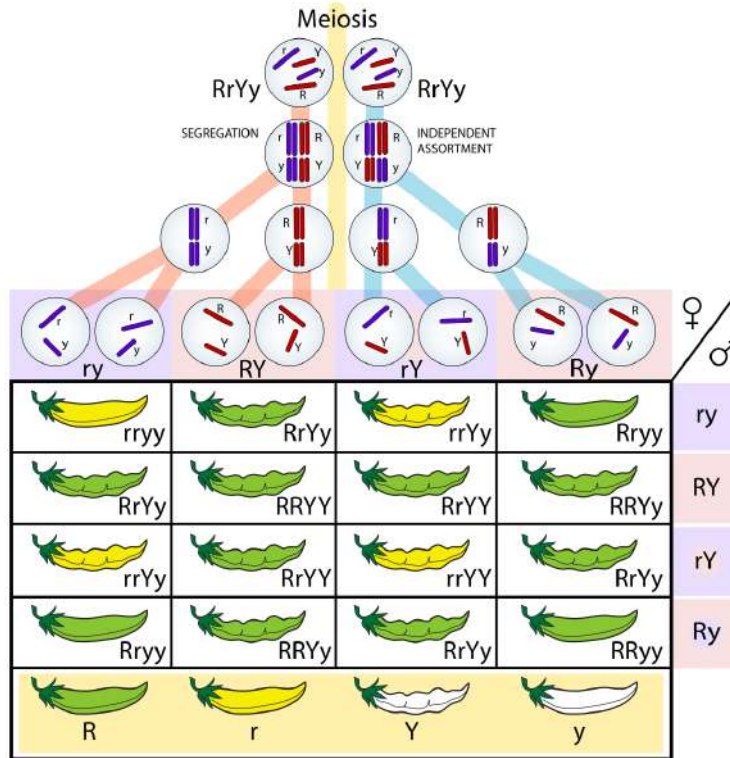


Fig. 1. Meiosis in pea plants depicting Mendel's Laws of Segregation and Independent Assortment. Refer to PowerPoint slides for color version. *Image credit: LadyofHats.*

We will come back to Gregor Mendel and genetic crosses in subsequent labs. First, we will need to understand basic probability theory and how it can be used to predict the likelihood of particular outcomes.

Part 1: Probability and Statistics

Probability can be defined as the chance that any particular outcome will occur. For example, what is the probability of tossing a coin and obtaining heads? The answer would be 1/2 or 50%. Thus,

$$Probability = \frac{\text{Number of times a particular event will occur}}{\text{Total number of possible outcomes}}$$



Going back to our coin-flipping example, we asked what the probability would be of flipping a heads on one try. Thus, the numerator (number of times a particular event will occur = 1) and the denominator = 2 (there are only two possible outcomes, heads or tails).

What would be the probability of drawing a black card from a deck of cards on one try? What would be the probability of drawing the King of Hearts from a deck of cards on the first try?

A) Random sampling error

When calculating probabilities, **random sampling error** can cause deviations from predicted probabilities. For example, if you tossed a coin six times you would predict that 50% of the tosses would be heads and 50% would be tails. However, it would be possible that you tossed heads twice and tails four times, leading to a high random sampling error and a deviation from the expected value of 50%. Conversely, if you tossed the same coin 1000 times it is highly likely that the number of heads and tails would be closer to 50%. Let's try this out in a few exercises.

Working in pairs, each group will obtain a deck of cards, a coin, and a dice. For each object, two tests will be conducted, one with a low sample number and one with a high sample number. This will enable us to determine the influence of random sampling error on our outcomes.

Coin test (10 flips)

	Expected	Observed	Difference
Heads			
Tails			

Coin test (100 flips)

	Expected	Observed	Difference
Heads			
Tails			

Cards test (20 cuts)

	Expected	Observed	Difference
Spades			
Clubs			
Hearts			
Diamonds			



Cards test (100 cuts)

	Expected	Observed	Difference
Spades			
Clubs			
Hearts			
Diamonds			

Dice test (30 rolls)

	Expected	Observed	Difference
One			
Two			
Three			
Four			
Five			
Six			

Dice test (120 rolls)

	Expected	Observed	Difference
One			
Two			
Three			
Four			
Five			
Six			

What can you conclude from these observations? How does sample size influence outcomes with respect to expectations? How do your results compare with other groups?

Although simply visualizing the results in a table can give a sense of how much the observed values deviate from the expected values, **statistical tests** can provide a more quantitative framework for hypothesis testing. Many statistical analyses require a **null hypothesis** that assumes no significant difference between treatments, events, or values. For example, in our experiment one null hypothesis would be that there is no significant difference between the expected and observed number of cards in each suit. We can test the null hypothesis using a statistical technique called a **chi square (χ^2) test**. In general, a low χ^2 is consistent with the null hypothesis, whereas a large χ^2 might lead us to refute the null hypothesis. How do you know if a value is large enough? First, let's see how we actually calculate χ^2 . The formula is relatively simple:



$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where

O = the observed data

E = the expected data based on the null hypothesis

Σ = summation over different categories

If we had two categories, the χ^2 would be calculated as

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

Once we have a χ^2 value we need to determine if it is statistically significant. This is accomplished using a χ^2 table (see table below). The values inside the table represent calculated χ^2 values. Before we can either reject or fail to reject the null hypothesis we need to determine our alpha (P-value) and degrees of freedom (df). P-values can be interpreted in multiple ways. For example, suppose we obtained a χ^2 of 0.016 with 1 df. Based on the table, random chance alone would produce a χ^2 value greater than 0.016 over 90% of the time (see bold square in table). In most statistical hypothesis testing we adopt an alpha (P-value) of 0.05 or 5%. Examining the table, with 1 df we would need a $\chi^2 > 3.841$ (known as the critical value) in order to reject the null hypothesis of no significant difference between observed and expected values. Degrees of freedom is obtained by subtracting 1 from the total number of categories (n - 1). For example, in our cards example df = 3.

Chi-Square Table

Table 5-2
Critical Values of the χ^2 Distribution

df \ α	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15



Working in groups of two, perform six χ^2 analyses using the data from the tables above. What can you conclude from your analysis?

B) The product rule

We can use an approach called the **product rule** to determine the probability of multiple events occurring. For example, what is the probability of flipping a coin five times and obtaining heads each time? To determine this, we can simply multiply the probability of each independent event occurring:

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/32 = 0.03125 = 3.125\%$$

What is the probability of drawing a diamond card three times in a row?

What is the probability of rolling a six 10 times in a row?

What is the probability of flipping heads 20 times in a row?

C) Binomial expansion

Note that the product rule is used to determine the probability of **ordered events**. It can be used to determine the probability of outcomes in succession. For example, what is the probability of drawing 10 red cards in a row from a standard deck of 52 cards? Conversely, the **binomial expansion equation** can be used to determine the probability of **unordered events**. For example, if a couple were to have seven children, what would be the probability that four are boys and three are girls? The binomial expansion equation is as follows:

$$P = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$



where P = the probability that the unordered outcome will occur
 n = the total number of events.
 x = the number of events in one category (e.g. number of males)
 p = individual probability of x .
 q = individual probability of other category.

The symbol '!' represents factorial. For example, $4! = 4 \times 3 \times 2 \times 1 = 24$.

In this example,
 $n = 7$
 $x = 4$
 $p = 0.5$
 $q = 0.5$

Plugging these numbers into our equation gives us the following:

$$P = \frac{7!}{4!(7-4)!} (1/2^4)(1/2^{7-4}) =$$

$$P = \frac{5040}{144} \left(\frac{1}{16}\right) \left(\frac{1}{8}\right) = 0.2734 \text{ or } 27.34\%$$

Binomial expansion problem: The disease cystic fibrosis is a recessive disease governed by a single gene. Only individuals homozygous recessive are affected, whereas heterozygous individuals are unaffected carriers. The disease causes difficulty breathing due to the buildup of mucous in the lungs. Suppose two heterozygous parents have five offspring. What is the probability that two of the five offspring will be affected with the disease?

To solve this problem it will be helpful to first create a Punnett Square depicting the possible genotypes of the offspring. Remember Mendel's Law of Segregation when constructing the Square and that both parents are heterozygous carriers of the cystic fibrosis allele.

Punnett Square



What would be the probability that the couple's first child is affected and the next four children are unaffected? Would you use the product rule or the binomial expansion formula to determine this?

Part 2: Measurements and dilutions

To adequately work in a genetics lab it is imperative to be comfortable with units of measurement and to be able to convert among units. Most of the time geneticists are working with liquids, so common units of measurement include milliliters (m) and microliters (μ). Below is a table illustrating the different notations.

Notation	Factor	Name	Symbol
10^{-1}	0.1	deci	d
10^{-2}	0.01	centi	c
10^{-3}	0.001	milli	m
10^{-6}	0.000001	micro	μ
10^{-9}	0.000000001	nano	n
10^{-12}	0.000000000001	pico	p
10^1	10	deka	da
10^2	100	hecto	h
10^3	1000	kilo	k
10^6	1000000	mega	M
10^9	1000000000	giga	G
10^{12}	1000000000000	tera	T

Using the table above, make the following unit conversions. Determine this by moving the decimal point the appropriate number of places.

1. 25 μ L = _____ mL
2. 3 L = _____ μ L
3. 150 cL = _____ dL
4. 5000 pL = _____ nL
5. 75 L = _____ kL
6. 800 nL = _____ cL
7. 6 hL = _____ ML
8. 8500 μ L = _____ L
9. 64 TL = _____ GL
10. 120000 nL = _____ μ L



In addition to being able to convert among different units of measurement, geneticists also need to make what is called a **working stock solution**. A working stock solution is a reagent or other solution that is directly used in reactions such as PCR, which we will study later on in the course. Most often, working stock solutions are made by **diluting an original stock solution** that is at a concentration higher than what is needed for the experiment. By concentration we mean the amount of solutes in a given volume. The concentration of solutes in a solution is usually measured by molarity (M), which is the number of moles per liter of solution (moles/L). For example, 1 M = 1 mole/L. To determine how many grams of a substance is equal to one mole the molecular mass is needed (we will not explore this in this lab). The general formula for making a dilution solution is the following:

$$C_1V_1 = C_2V_2$$

where

C_1 = the original concentration of stock solution

V_1 = the volume of stock solution needed to make working solution

C_2 = the desired concentration of working solution

V_2 = the desired final volume of working solution

How would you make a 100 μ L working stock at a concentration of 0.5 μ M from a stock solution at a concentration of 0.1 mM? Always remember to put everything into the same units first!

What would be the working stock concentration (in molarity) if you diluted 250 μ L of a stock concentration of 10 μ M in a total volume of 1 mL?

Part 3: Biological Controls

In most scientific experiments, a **control** is needed to make sure that the experiment was successful and to guard against incorrect conclusions. Controls are particularly needed in the field of **molecular genetics**. As we will learn later in the semester, a technique called polymerase chain reaction (PCR) can be used to amplify (i.e. make many copies of) a particular segment of DNA. The following components are needed to perform PCR:

1. Water
2. PCR buffer
3. dNTP (nucleotides)
4. Primers (oligonucleotides) that flank the region of DNA to be amplified
5. Taq polymerase enzyme
6. DNA



We will discuss how PCR is actually performed later in the semester. For now, let's think about why controls are needed when we perform these experiments. First, it's important to understand what we mean by a **negative control**. In the case of molecular genetics and PCR, a negative control is a sample that we add to our experiment that should **not produce a positive reaction**. If template DNA is required for successful amplification, we would not expect to see a product from a sample lacking any DNA. Figure 2 below shows a gel image resulting from a PCR with seven real samples (with DNA) and one negative control (water). Note that PCR product (black bands) can be seen for each of the samples, but not the negative control (water). This would be considered a successful run.

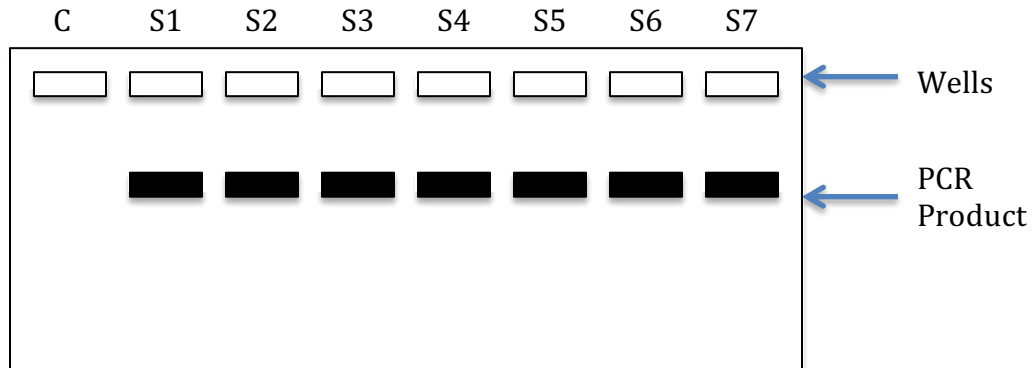


Fig. 2. Gel image illustrating the concept of a negative control. Well 'C' contained all of the necessary PCR reagents, but substituted water for DNA. Thus, we would not expect to see any PCR product from this sample, which is confirmed by the image.

So why is the inclusion of a negative control useful? Suppose that a band was also present in our water sample. This would indicate the presence of DNA contamination of our negative control (water). Since this happened, there might also be contamination of our real samples and therefore the results may be suspect. In general, contamination of the negative control requires that the experiment be conducted again, using a new negative control and possibly new reagents (e.g. water, buffer, primers, dNTPs).

Although not used as frequently as negative controls, some experiments may make use of **positive controls**. Positive controls are samples that we know should work/amplify using a given set of experimental conditions. If the positive control works, but the samples do not, this would indicate an issue with the samples. If neither the samples nor the positive control amplifies, this may indicate an issue with the conditions and/or reagents used in the experiment. We will make use of both negative and positive controls later on in the semester when we work on the GMO activity.

Cytogenetics and Karyotyping

Overview and objectives

This week we will be learning how cytogeneticists use karyotyping to understand chromosomal abnormalities. **Karyotyping** is a technique where chromosomes are stained and visualized during the metaphase stage of cell division (mitosis or meiosis). In this lab, you will take on the role of a medical cytogeneticist and use human karyotypes to diagnose various diseases and abnormalities in patients. Upon completion of the lab, you should be familiar with what chromosomes are, what a karyotype is and how it is constructed, the different ways in which chromosomal abnormalities might arise, and the basic terminology used to describe chromosomes.

Background information

Overview

DNA is packed into units called **chromosomes**. In many species, DNA is closely associated with several types of proteins called histones that are used to tightly pack the DNA into the cell nucleus (in eukaryotes). In eukaryotes, we say that the chromosomes are linear, whereas most bacteria contain a single circular chromosome. All eukaryotes also contain a second genome inside their cells called mitochondrial DNA (mtDNA). Genes in mtDNA help regulate the process of cellular respiration. In addition, plants and several protists contain a third genome in their chloroplasts called chloroplast DNA (cpDNA), which is involved with photosynthesis. In this lab we will focus exclusively on **eukaryotic nuclear chromosomes**. Human somatic cells contain 46 chromosomes. As humans are a diploid species, half of the complement of chromosomes (23) originated from a sperm cell and half (23) originated from the egg cell. These are called **homologous chromosomes**. For example, each person has two copies of Chromosome 1, one copy inherited from the father and one copy inherited from the mother. Portions of homologous chromosomes can exchange segments during meiosis in a process called recombination. During cell division, each homolog is also duplicated, forming sister chromatids (Fig. 1). The region where sister chromatids attach during cell division is called the **centromere**. **Telomeres** are regions at the tips of chromosomes that consist of highly repetitive sequences that form a protective cap to the ends of chromosomes. Telomeres tend to shorten with each cell division, leading to cell aging and eventually cell death. Much research has focused on techniques decrease the propensity of telomere shortening to counteract the aging process. Therapeutic-related work has also focused on the role of telomeres in the proliferation of cancer cells.

Human sex is determined genetically through the XY system—human females are XX and human males are XY. Research has shown that human “maleness” is determined by a gene that sits on the Y chromosome. In several other species, sex is not determined by genes and chromosomes, but by environmental temperature through what is called temperature-dependent sex determination. For example, in many crocodylian species, males are only produced if eggs are incubated at intermediate temperatures. In contrast, females are produced if eggs are incubated at either extreme.

There is a large variation in the number of chromosomes among species, but there tends to be a correlation between organismal complexity and genome size. The human genome is approximately 3 Gbp, or 3 billion base pairs long. This is considered a relatively large genome (some salamander genomes are much larger), but surprisingly only about 2% of the genome encodes a functional product (e.g. proteins). Many researchers refer to the remaining 98% as ‘junk DNA’, although it is likely that other regions of the genome are involved in regulating gene expression to some degree. For such a large genome to fit inside the cell nucleus, it needs to be condensed substantially. The DNA is first wrapped around histone proteins, which then associate to form what are called nucleosomes. The nucleosomes further compress into what is called a 30 nm fiber. Additional coiling eventually gives chromosomes their highly compact, typical appearance during cell division.

Chromosome classification

Chromosomes are generally classified using multiple criteria. First, they are numbered from largest to smallest. For example, human chromosome 1 would be the largest chromosome, which contains 2,100 protein-coding genes and 249 million bp. The sex chromosomes are labeled appropriately as either XX in females or XY in males. Note that the X-chromosome is much larger than the Y-chromosome. Therefore, the X- and Y- chromosomes are considered non-homologous (although there are a few homologous regions that are needed for proper pairing during cell division).

The relative position of the centromere can also differ between chromosomes. Centromeres that are placed in the center of a chromosome are called **metacentric chromosomes**, resulting in equal length chromosomal arms. When the centromere is not found in a central position, different chromosomal arm lengths result. These arms are referred to as **p arms (short arms)** and **q arms (long arms)**. In addition to metacentric chromosomes, chromosomes can be **submetacentric, acrocentric, and telocentric**, all varying in the relative position of the centromere (Fig. 2).

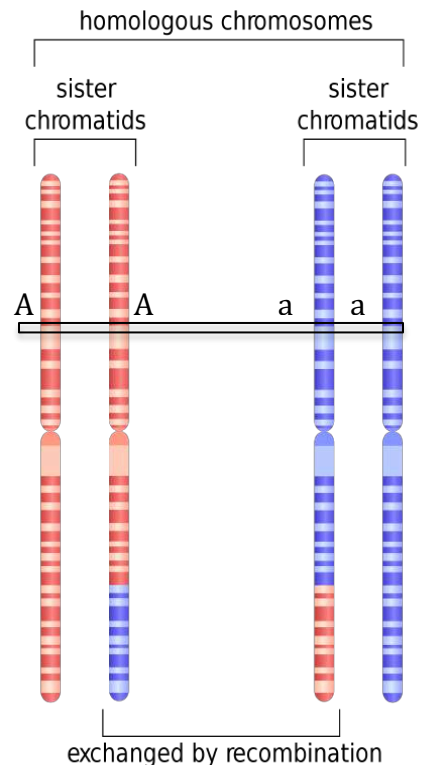


Fig. 1. Homologous chromosomes and sister chromatids. Note that recombination has taken place. Box and letters illustrate heterozygosity at the ‘A’ gene. *Image credit: CC BY-SA 3.0 Emw (author).*

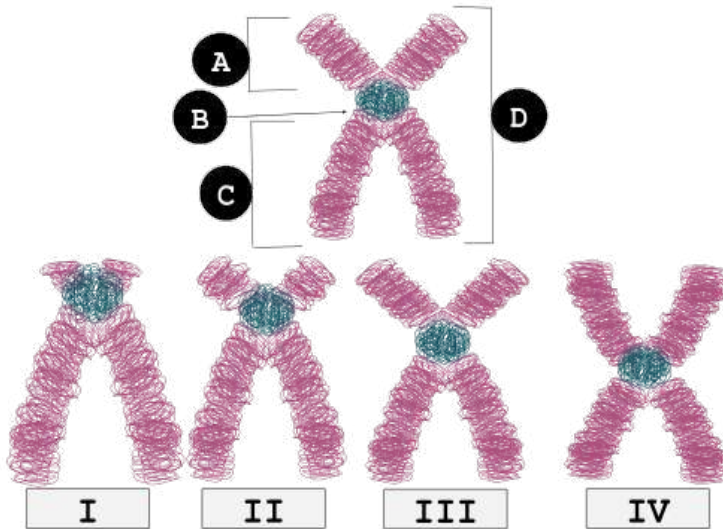


Fig. 2. Alternative means of classifying chromosomes based on the relative position of the centromere. I = telocentric; II = acrocentric; III = submetacentric; IV = metacentric. A = p arm; B = centromere; C = q arm; D = sister chromatid. *Image credit: CC BY-SA 4.0 Fockey003 (author).*

A final way that cytogeneticists can differentiate and classify chromosomes is to stain them with dyes that produce diagnostic banding patterns. Note that bands on a stained chromosome are not indicative of genes. In general, there can be dozens or hundreds of genes within a single band. A common stain used by cytogeneticists is called the **Giemsa stain**, also known as the G-banding technique. Following the staining procedure, AT-rich regions of the chromosome appear dark whereas transcriptionally active GC-rich regions incorporate less of the dye and thus appear lighter. The differential banding patterns can be used to identify homologous chromosomes and look for chromosomal alterations.

Procedure for creating karyotype

1. Obtain cells (e.g. white blood cells, skin cells, cancer cells, cells from amniotic fluid). Obtaining cells from amniotic fluid (amniocentesis) can be used to detect genetic abnormalities in the fetus.
2. Treat cells with chemical to promote cell division.
3. Treat cells with a second chemical that halts cell division at metaphase. This is the stage when chromosomes are highly condensed and easily visualized.
4. Immerse cells in a hypotonic solution and place on glass slide.
5. Stain chromosomes with Giemsa stain and observe under microscope.
6. Prepare digital images and analyze results.

Genetic abnormalities observed by G-banded karyotyping

In general, chromosome abnormalities can include both **numerical changes** and **structural changes**. Numerical changes can involve an irregular number of a particular chromosome (**aneuploidy**). For example, a diploid individual possessing three copies of a chromosome instead of the usual two copies would be called **trisomy**, whereas an individual with one copy would be called **monosomy**. Trisomy 21 would mean three copies of chromosome 21 and is characteristic of individuals with Down syndrome (Fig. 3).

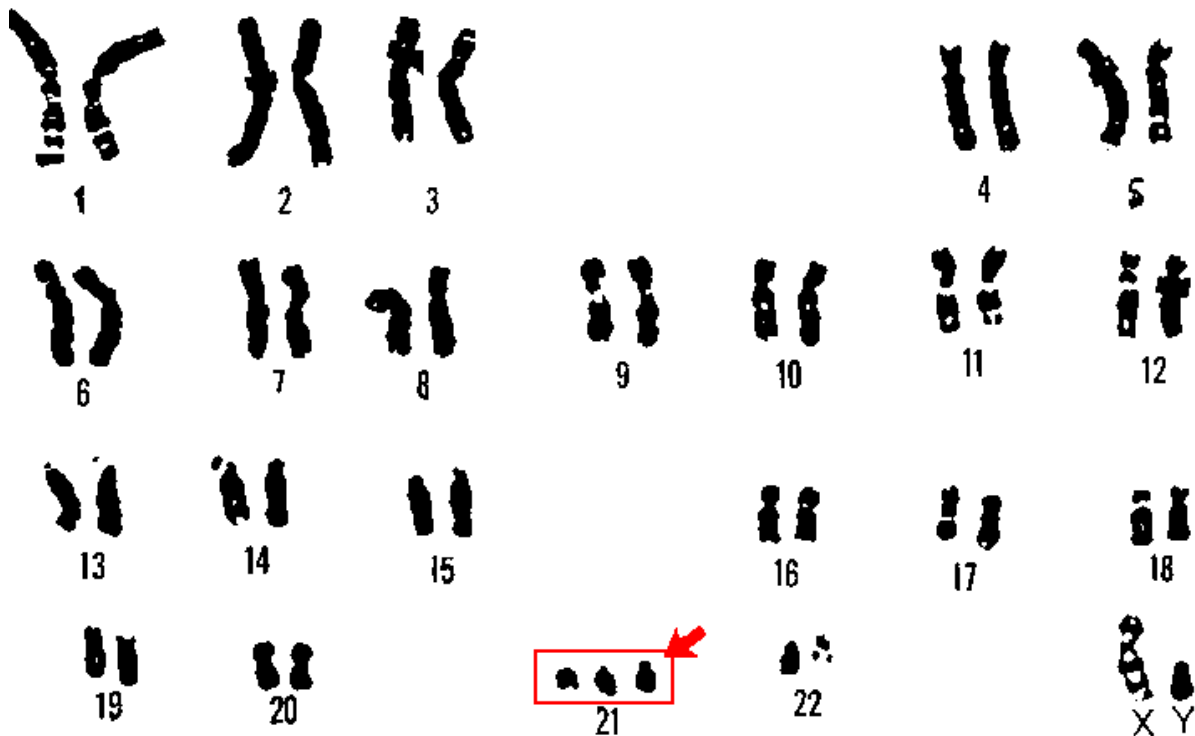


Fig. 3. Karyotype of a human male with Trisomy 21 (Down syndrome). Note the three copies of Chromosome 21. Homologous chromosomes were determined by size and banding pattern. *Image credit: US Department of Energy Human Genome Program.*

In addition to abnormalities in the number of specific chromosomes, an individual can have an incorrect number of the complete set of chromosomes. A gain in one or more complete sets of chromosomes is termed **euploidy**, whereas a loss of an entire set of chromosomes is called **monoploidy**. Monoploidy is usually fatal due to recessive mutations (many of which are lethal) being phenotypically expressed.

In addition to numerical changes, chromosomes can also exhibit the following structural changes involving either one or two chromosomes (Fig. 4):

- a. Translocations (two chromosomal change) – two non-homologous chromosomes fuse or swap segments.
- b. Insertions (two chromosomal change) – one region of one chromosome removed and inserted into a different chromosome.
- c. Inversions (single chromosomal change) – certain regions of a chromosome are placed in an incorrect orientation with respect to the remainder of the chromosome.
- d. Duplications (single chromosomal change) – one segment of a chromosome is duplicated.

- e. Deletions (single chromosomal change) – removal or loss of a particular region or segment of a chromosome. Commonly occurs in telomeric regions during cell division.

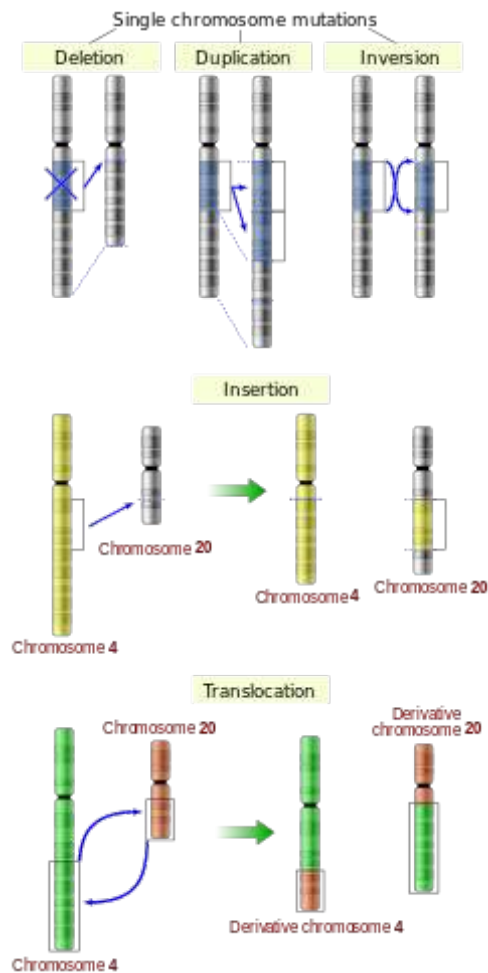


Fig. 4. Alternative mechanisms that can contribute to structural chromosomal abnormalities. Note that insertions and translocations involve more than one chromosome. *Image credit: YassineMrabet (author)*

Although standard karyotyping can provide a powerful approach to diagnosing various genetic diseases, it is considered a wide resolution technique, meaning that it is unable to detect genetic abnormalities below a certain number of base pairs (~1 million). In these cases, other techniques such as fluorescent in situ hybridization (FISH), array comparative genome hybridization (CGH), and direct DNA sequencing can be used, the latter of which can be used to detect single nucleotide changes or polymorphisms (SNPs).

Activity

Working in pairs, you will assume the role of a medical cytogeneticist and aid in the diagnosis of various genetic diseases. Each group should try to get through as many case studies as possible.

1. Each group should receive several blank copies of the Cytogenetics Report to be filled out for each case study.
2. Each Chromoscan Board contains a single case study, is **color-coded** and is associated with the colored chromosome decals. For example, Case Study A is pinkish purple. **Never mix and match Chromoscan Boards and decals.**
3. Randomly select one of the chromosome decals and make a sketch of the chromosome on the Cytogenetics Report. Be sure to **label the centromere, telomere, p arm, and q arm**. Also make note if the chromosome is **metacentric, submetacentric, acrocentric, or telocentric**.
4. Record all the relevant information from the case study on the report.
5. To create the karyotype you will match the correctly colored chromosome decals with the chromosomes already present on the Chromoscan Board. In other words, your goal is to correctly pair homologous chromosomes.
6. Once the karyotype is created, analyze it for any chromosomal abnormalities (e.g. inversions, translocations, deletions, insertions, duplications).
7. Report your findings on the Cytogenetics Report.
8. To diagnose the patient, refer to the table below and write the diagnosis on the Cytogenetics Report.
9. Using your smartphones, provide additional clinical information (in addition to what is presented in the case study) about the disease on the Cytogenetics Report.
10. Remove chromosome decals and randomly place them in the cryostorage region of the Chromoscan Board.
11. Repeat procedure for additional case studies. **Do not share answers with other groups!**



Diagnostic Table

Karyotype Findings	Diagnosis
No abnormalities found	None
Trisomy 21	Down syndrome
XXX	Trisomy X female
Inversion on chromosome 3 involving p arm, centromere and q arm	Not associated with any known disease
Trisomy 18	Edwards syndrome
Trisomy 13	Patau syndrome
Reciprocal translocation with chromosomes 9/22	Chronic myelogenous leukemia
Robertsonian translocation (fusion of chromosome arms) with chromosomes 14/21	Down syndrome
XXY	Klinefelter syndrome
XYY	XYY male
5p deletion	cri du chat
Monosomy X	Turner syndrome
Inversion on chromosome 9	Not associated with any known disease

Cytogenetics Report

Names: _____ Date _____

Directions: Each group should complete a separate Cytogenetics Report **for each case study (print out several sheets)**. After a case study is completed, place the chromosome decals randomly in the cryostorage regions of each board. Do not mix and match chromosomes and boards. Before assembling the karyotype, randomly select one chromosome and make a detailed sketch in the space below. Note the location of the **centromere, telomeres, the p arm, and the q arm**. Also note if the chromosome is metacentric, submetacentric, acrocentric or telocentric.

Chr. type: _____metacentric _____submetacentric _____acrocentric _____telocentric

Patient name	Case Study ID	Patient Age
Why is the patient being referred for karyotyping?	Source of cells: _____Blood _____Amniocytes _____Chorionic Villi _____Other	
Total number of chromosomes observed	Gender	
Chromosomal Findings _____no abnormalities _____monosomy (chr _____) _____trisomy (chr _____) _____deletion (chr _____, arm _____) _____insertion (chr_____, arm_____) _____translocation (chrs____and____) _____inversion (chr____, arm____)	Patient diagnosis and additional information to provide to patient and family.	



Preparation of materials for Lab 3

Your final objective this week is to prepare fungal genetic crosses to observe the following week. This experiment provides a simple, yet effective hands-on introduction to basic properties of **meiosis**. Our technicians have already prepared cultures of three types of *Sordaria fimicola* (**wild-type, mutant tan, and mutant gray**). In the lab you should see six dishes, two containing the wild-type strain, two with the tan strain, and two with the gray strain. Your job is to set-up crosses of these different strains to determine the proportion of asci containing “hybrid” ascospores that are a result of crossing over during prophase I of meiosis. You will have multiple dishes containing agar that will be used to perform your experiment.

1. Use alcohol to disinfect all work surfaces and wash your hands.
2. Obtain a petri dish containing the *Sordaria* crossing agar to make your crosses. Remember that we are making crosses between wild-type and gray and wild-type and tan strains. For simplicity, I recommend using separate plates for the tan and gray mutants. For example, one dish will be a wild-type vs. tan cross and another dish will contain wild-type vs. gray cross.
3. Using a marking pencil, invert your dish and draw four equal quadrants. Label two of the quadrants as wild-type and two as mutant (either gray or tan). Refer to Fig. 5 for an example. Also, make sure to write your names and the date on your dishes.

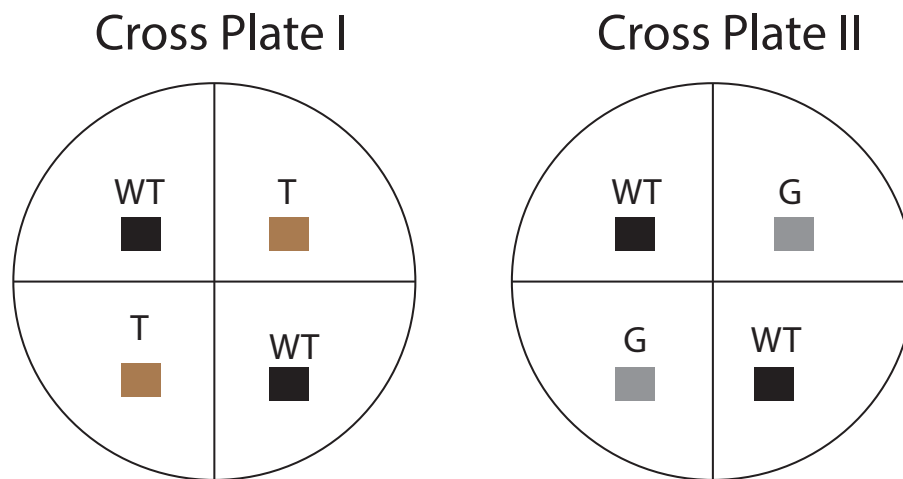


Fig. 5. Creation of *Sordaria* crosses. Cross Plate I illustrates cross between wild-type black and mutant tan strains. Cross Plate II shows cross between wild-type black and mutant gray strains.

4. Using a **sterilized wooden splint**, cut small 0.5 cm cubes from the stock culture dishes. It is important to use sterilized equipment to minimize contamination. The common procedure for sterilization is to dip a utensil (e.g. scalpel) into alcohol and then flame it over a Bunsen burner. In our case, the wooden splits are already sterilized. Carefully invert the cubes (hyphae side down) and place them in the correct location on the cross plates (see diagram above). For example, you will cut out a 0.5 cm cube from the tan strain and place it in one of the quadrants labeled “tan” on your cross plate. Each culture plate should contain four cubes (two wild-type and two mutant [either tan or gray, but not both]). Using the side of your splint, gently press the cubes onto the agar so they stick. Next, apply some parafilm and seal the outside of your dishes.
5. Incubate the dishes at room temperature for 1 week.



Meiosis and Genetic Recombination

Background:

Overview of Meiosis: In this lab we will examine cell division by meiosis. Meiosis, unlike mitosis, results in a change in ploidy among daughter cells. In diploid species such as humans, meiosis results in four (4) daughter cells that originate from a single diploid germ cell. In humans, all somatic (non sex cells) are diploid ($2n$) and contain two sets of chromosomes ($2n = 46$). Haploid cells ($n = 23$) are found only in gametes (sperm and egg). During the process of fertilization, the haploid genome of the sperm fuses with the haploid genome of the egg to return to the diploid state in the resulting zygote.

Also, unlike mitosis, meiosis is divided into two separate divisions termed **meiosis I** and **meiosis II**. **Homologous chromosomes** separate in meiosis I, whereas **sister chromatids** separate in meiosis II. Figure 1 illustrates both stages of meiosis using an example with $2n = 6$. Unlike interphase, chromosomes are highly condensed during meiosis, which makes them clearly visible using standard light microscopy. Looking at metaphase I we can see that homologous chromosomes pair up at the metaphasic plate. For example, the copy of chromosome 8 inherited from the father aligns with the copy of chromosome 8 inherited from the mother. This contrasts with metaphase in mitosis, where homologous chromosomes simply line up in a single line one chromosome wide. The pairing and random organization of homologous chromosomes along the metaphasic plate during meiosis I is one mechanism that leads to daughter cells being genetically different from the parent cell. We can call meiosis the 'reductional phase' because it produces two haploid cells from a single diploid cell. Meiosis II is similar to mitosis in the separation of sister chromatids. Also, as in mitosis, during meiosis II chromosomes align into a single column during metaphase II. **Take some time to make sure you understand Fig. 1 and the stages of meiosis, as it will be important for the lab activity.**

If you look closely at Fig. 1, you will notice that some chromosomes contain both blue and red segments. This is due to a process called 'crossing over' or recombination and occurs during prophase I. During crossing over, homologous portions of two chromosomes swap information, leading to chromosomes that contain a mix of alleles from both maternal and paternal chromosomes. For many species, crossing over is facilitated by the formation of a **synaptonemal complex**, which helps bind the arms of homologous chromosomes. By examining the rate of recombination among alleles, researchers can construct a **linkage map** that shows the position of genes relative to one another along a chromosome. This is because the probability of recombination is proportional to the distance separating two genes along a chromosome. In other words, alleles at two genes that are physically close together on a chromosome are more likely to be inherited together versus alleles separated by a large distance. Let's take a look at another example in Fig. 2.

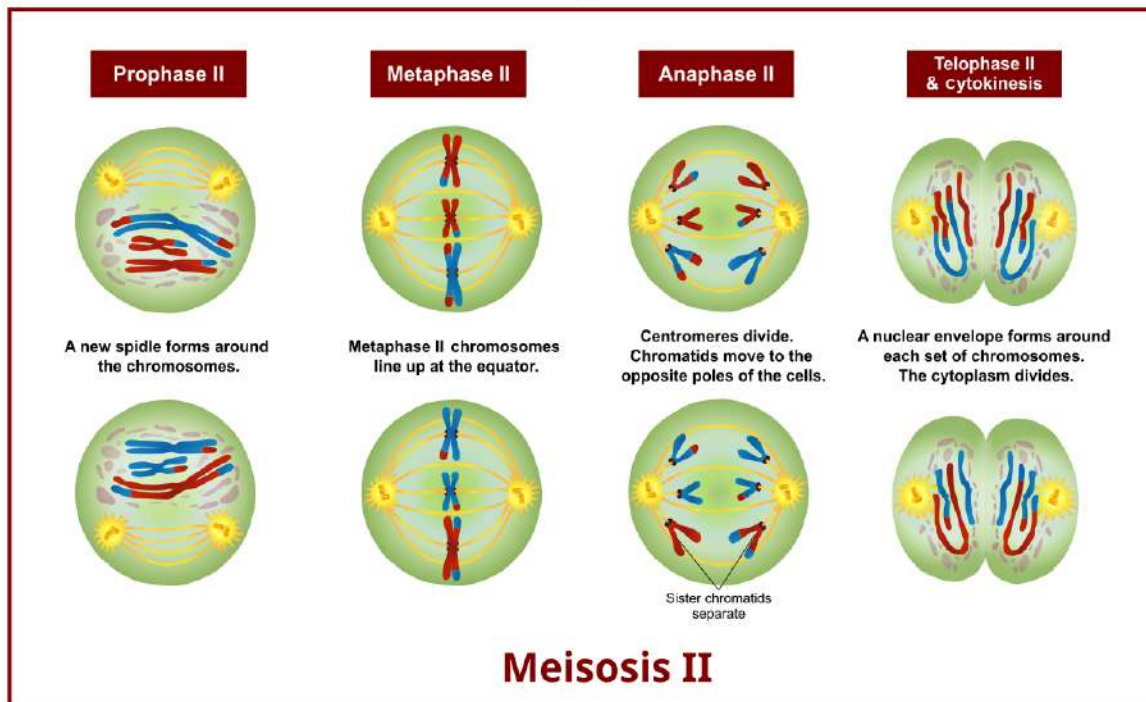
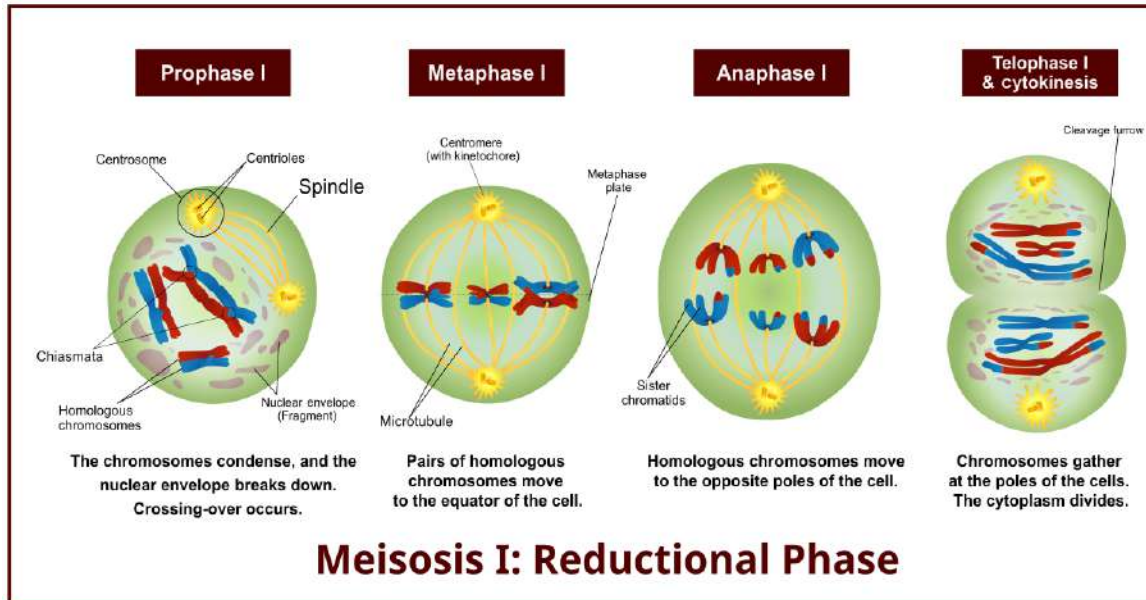


Fig. 1. Overview of meiosis in a cell with 6 chromosomes ($2n = 6$). For simplicity, paternal chromosomes are in blue and maternal chromosomes are in red. Note that upon completion of meiosis II, four haploid cells are produced. The yellow circles at the poles are centrosomes, which anchor the mitotic spindle where chromosomes attach via their centromere. Proteins called the kinetochore help attach centrosomes to spindle. *Image credit: CC BY-SA 4.0 Ali Zifan (author).*

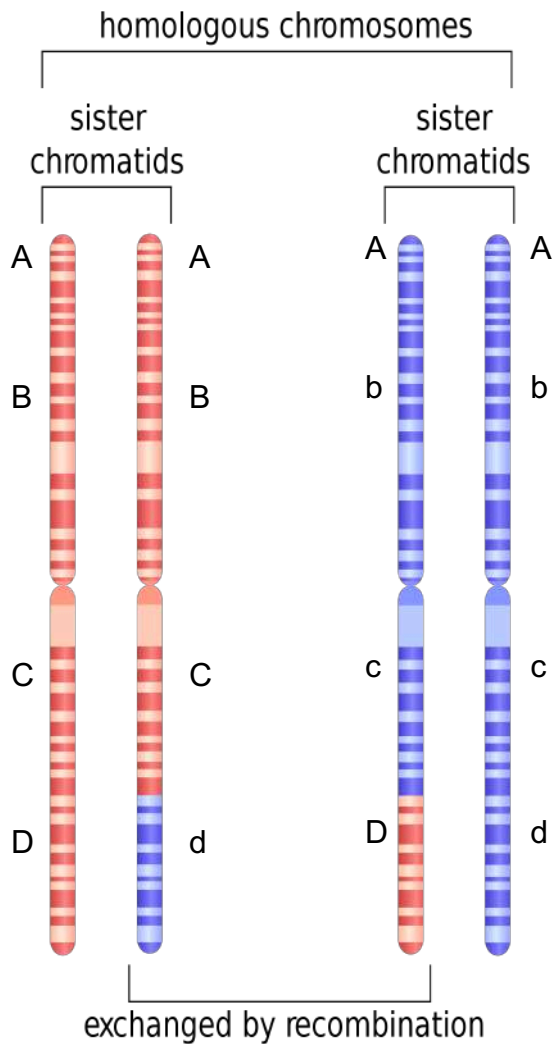


Fig. 2. Example of genetic recombination involving two homologous chromosomes, each consisting of two identical sister chromatids. Letters denote the location of genes and alleles on each chromosome. *Image credit: CC BY-SA 3.0 Emw (author).*

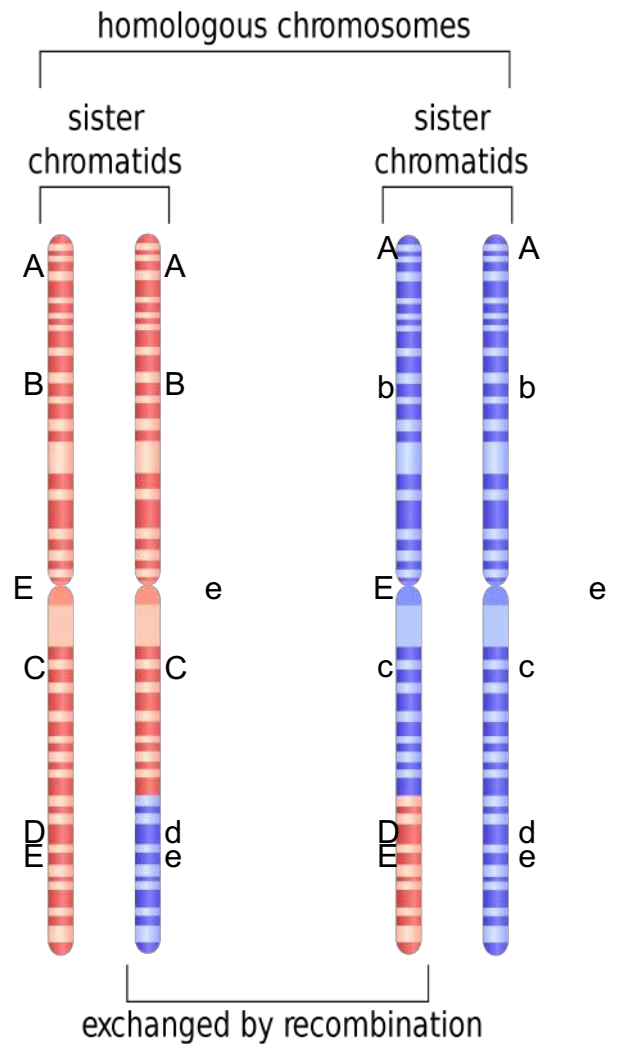


Fig. 3. Example of genetic recombination involving two homologous chromosomes, each consisting of two identical sister chromatids. Letters denote the location of genes and alleles on each chromosome. *Image credit: CC BY-SA 3.0 Emw (author).*

Here, we are working with four genes we will call Gene A, Gene B, Gene C, Gene D. This particular individual is homozygous for Gene A, heterozygous for Gene B, heterozygous for Gene C, and heterozygous for Gene D. The maternal chromosome originally contained the alleles A, B, C, and D, whereas the paternal chromosome contained alleles A, b, c, and d. Following recombination, the d allele becomes associated with the A, B, and C alleles on one maternal chromatid. Further, the D allele becomes associated with the A, b, and c alleles on one of the paternal chromatids. Thus, there is a new 'combination' of alleles that will be transmitted to daughter cells and eventually to sperm or egg.

In Fig. 3 we can add another gene, Gene E that is in close proximity to Gene D. Because these two genes are so close to one another, the likelihood that recombination will occur between them is minimal. Thus, alleles present at both genes will tend to be linked, or always inherited together. For example, the D and E alleles will be inherited and the d and e alleles will be inherited (never D and e or e and D). Because these alleles tend to be present in particular combinations in offspring, we can deduce that the genes must be in close proximity on the chromosome.

Biology of Sordaria fimicola: In this lab we will be working with cultures of the fungus *Sordaria fimicola*. This is a species of fungus that is commonly used in genetics labs to illustrate properties of meiosis and recombination. The fungus is commonly found in the dung of herbivorous animals. *Sordaria* belong to a phylum of fungi called Ascomycota (sac fungi). Reproduction occurs sexually via the production of **haploid spores called ascospores**. These ascospores are contained in a structure inside the fungus called an **ascus**. All of these structures are contained in a **perithecium**. Each ascus contains eight ascospores that get released into the environment. Wild-type spores are a dark color, whereas mutant varieties are tan or gray. The full life cycle of a typical ascomycete fungus is shown in Fig. 4.



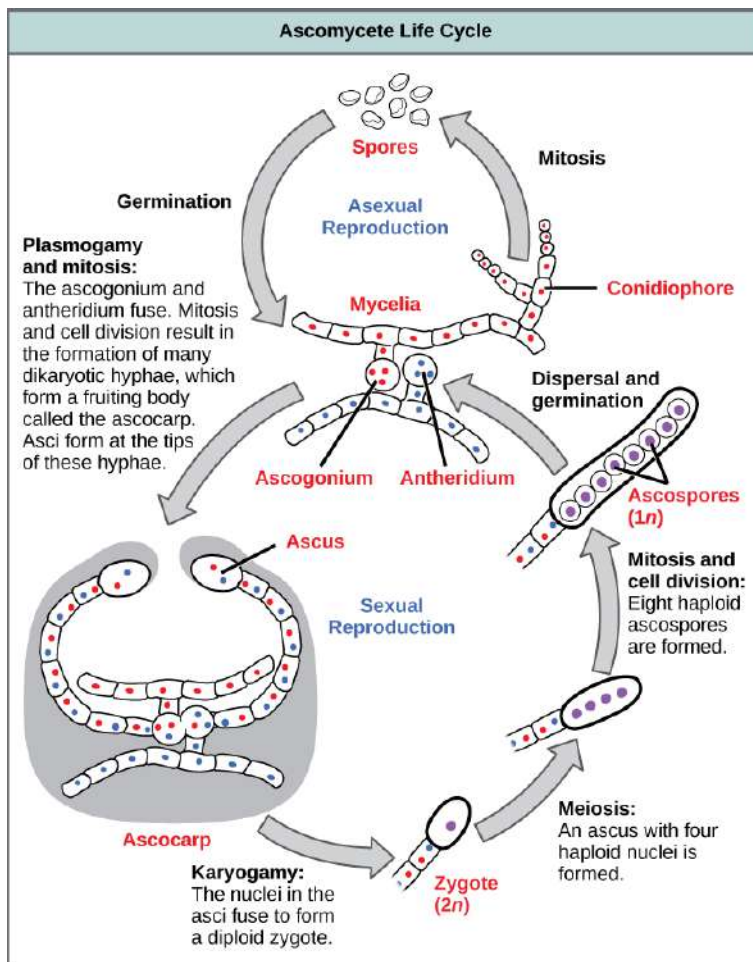


Fig. 4. Sexual and asexual reproductive cycles of a typical ascomycete fungus. Colored dots inside cells represent nuclei. Note that most cells of the fungus exist in a haploid or a dikaryotic stage. *Image credit: CC BY-SA 4.0 Zeeshan 93 (author).*

Focusing our attention on the sexual life cycle, we start with the fusion of **hyphae** from different mycelia. Following mitosis, this results in the production of a **fruiting body called an ascocarp**. Inside the ascocarp are many dikaryotic cells containing unfused nuclei. Upon fusion of nuclei (**karyogamy**) in the ascus, a diploid zygote is produced. This zygote undergoes meiosis to produce four haploid nuclei that subsequently undergo mitosis to produce eight haploid ascospores. These spores get released into the environment and germinate into a new haploid mycelium.

If no recombination (crossing over) has occurred for genes controlling ascospore color, we anticipate an **ordered 4:4 ratio of wild-type to mutant spore color** following a cross between strains (Fig. 5). Conversely, if crossing over did occur, we would expect deviations from the 4:4 ratio (e.g. 2:2:2:2 or 2:4:2).

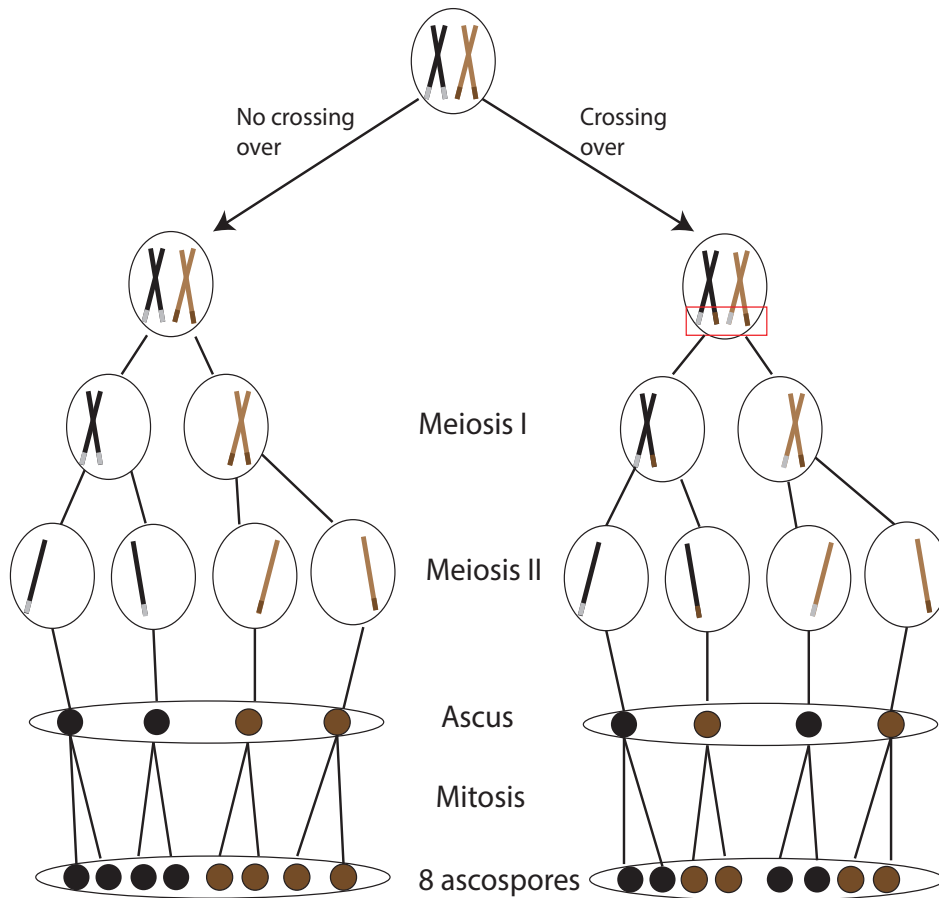


Fig. 5. Graphical depiction of the formation of ascospores in *Sordaria fimicola* following a cross between a wild-type black and a mutant tan strain. The left path depicts a situation where no recombination occurs among genes controlling ascospore color, producing the typical 4:4 ratio of black to tan. Conversely, the path on the right shows a case of recombination. In this case, recombination produces ascospores in a ratio of 2 black:2 tan:2 black:2 tan. The gray color on the black chromosomes represent the location of genes governing black spores, whereas the dark brown color on tan chromosomes indicate the location of genes controlling tan spore color. Red box highlights crossing over event.

The goal of this laboratory is to use cultured *Sordaria* to better understand meiosis and genetic recombination, and how meiosis can lead to offspring that differ genetically from parents.

Week 2 – Preparation of crosses (this is what was completed the previous week):

Our technicians have already prepared cultures of three types of *Sordaria* (wild-type, mutant tan, and mutant gray). In the lab you should see six dishes, two containing the wild-type strain,

two with the tan strain, and two with the gray strain. Your job is to set-up crosses of these different strains to determine the proportion of asci containing “hybrid” ascospores that are a result of crossing over during prophase I of meiosis. You will have multiple dishes containing agar that will be used to perform your experiment.

1. Use alcohol to disinfect all work surfaces and wash your hands.
2. Obtain a petri dish containing the *Sordaria* crossing agar to make your crosses. Remember that we are making crosses between wild-type and gray and wild-type and tan strains. For simplicity, I recommend using separate plates for the tan and gray mutants. For example, one dish will be a wild-type vs. tan cross and another dish will contain wild-type vs. gray cross.
3. Using a marking pencil, invert your dish and draw four equal quadrants. Label two of the quadrants as wild-type and two as mutant (either gray or tan). Refer to Fig. 6 for an example. Also, make sure to write your names and the date on your dishes.

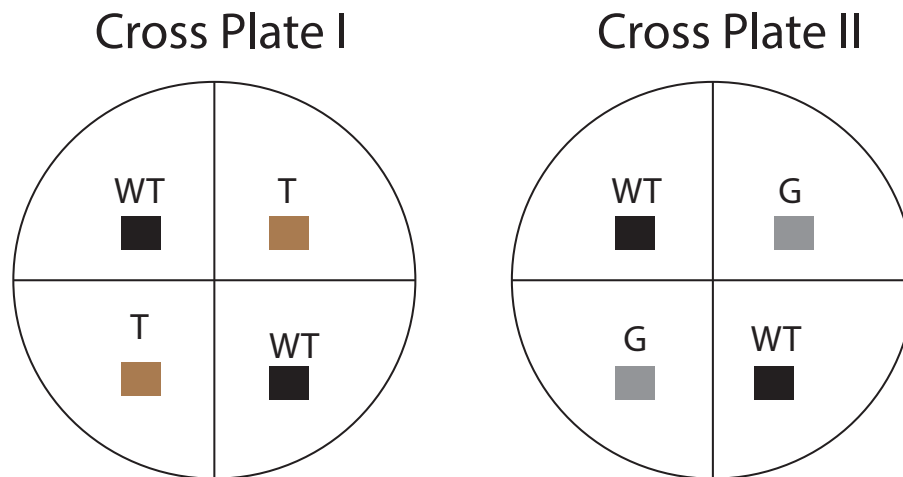


Fig. 6. Creation of *Sordaria* crosses. Cross Plate I illustrates cross between wild-type black and mutant tan strains. Cross Plate II shows cross between wild-type black and mutant gray strains.

4. Using a **sterilized wooden splint**, cut small 0.5 cm cubes from the stock culture dishes. It is important to sterilize all equipment to minimize contamination. The common procedure for sterilization is to dip a utensil (e.g. scalpel) into alcohol and then flame it over a Bunsen burner. In our case, the wooden splints are already sterilized. Carefully invert the cubes (hyphae side down) and place them in the correct location on the cross plates. For example, you will cut out a 0.5 cm cube from the tan strain and place it in one

of the quadrants labeled “tan” on your cross plate. Each culture plate should contain four cubes (two wild-type and two mutant [either tan or gray, but not both]). Using the side of your splint, gently press the cubes onto the agar so they stick. Next, apply some parafilm and seal the outside of your dishes.

5. Incubate the dishes at room temperature for 1 week.

Week 3 – Microscopic examination of crosses:

This week we will examine the results of the genetic crosses we made between wild-type and mutant *Sordaria* strains during the previous week. Your goal is to determine if asci show signs of genetic recombination, based on the expected distribution of spore color. **Remember, we expect a 4:4 wild-type to mutant distribution in asci that did not undergo crossing over.** Follow the instructions below to visualize the asci.

1. Use alcohol to disinfect all work surfaces and wash your hands.
2. Working in groups of two, obtain a compound light microscope, glass slides, cover slips, water dropping bottles, and sterilized toothpicks.
3. Using a sterilized tooth pick, remove a few perithecia from the cross plate and create a wet mount. **You can find the perithecia near the junction of the different strains.** To create a wet mount, gently rub the tip of the toothpick on a glass slide to remove the perithecia. Next, place a drop of water on your specimen.
4. Using a cover slip, gently press down on the perithecia until they rupture and release the asci. Note that putting too much pressure will force the ascospores out of the ascus, making it impossible to draw any conclusions from the experiment.
5. Observe the asci and ascospores under low magnification. Remember that wild-type spores should be a dark color, whereas the two mutant strains will be tan or gray. Next, try switching to high power magnification to view the ascospores more closely. Remember, crossing over has occurred in asci demonstrating either a 2:2:2:2 or 2:4:2 color pattern. Refer back to Fig. 5 to review how and when crossing over takes place.
6. Each group should count **200 asci**, keeping track of how many show evidence for recombination and how many do not.
7. We can actually use the frequency of crossing over to estimate the relative position of the genes on the chromosome with respect to the centromere. This is because there is a direct relationship between recombination rate and distance between two points (genes) along a chromosome. Fill out the table below with your data.



Strains Crossed	# Non-recombinant asci	# Recombinant asci	Total asci	% Recombinant	Map Units (% Recombinant/2)
Wild-type x gray					
Wild-type x tan					

Review questions:

1. Compare and contrast mitosis from meiosis. In your answer, make sure you address why mitosis generally results in two genetically identical daughter cells, whereas meiosis leads to genetic variation. Feel free to use drawings in your answer.

2. Meiosis is involved in the process of sexual reproduction. Sexual reproduction can lead to offspring that differ genetically from parents. What are the benefits of sexual reproduction and increased genetic diversity? What are the potential costs of sexual reproduction? Can you think of a real-world example that might favor asexual reproduction?

3. In your own words, describe the typical sexual life cycle of *Sordaria*. Make note of which stages are haploid, diploid, and dikaryotic. Are ascospores produced by meiosis or mitosis? Explain your reasoning.



4. Based on what you know of mitosis and meiosis, complete the following sentence:
Crossing over occurs during the _____ stage of _____.

5. Discuss the relationship between recombination rate and the position of genes along a chromosome. Would a high recombination rate indicate that two genes are close together or far apart along a chromosome? Explain your answer.



Patterns of Inheritance: Fruit Flies, Blood Types, Population Genetics and Hardy-Weinberg

Lab Overview

Over the next four weeks we will be working with two different exercises and simulations to further examine patterns of inheritance. In the first series of labs you will perform monohybrid and dihybrid crosses using real fruit flies (*Drosophila melanogaster*) to examine inheritance patterns. The labs are spread out over a course of four weeks due to the time it takes to cross and rear flies. The second set of labs deals with inheritance patterns of human blood types. During the “down time” of the fly lab, you are expected to work on the blood typing activity and become familiar with the genetics behind different blood types. You are also expected to become familiar with the **Hardy-Weinberg principle**, and how it can be used as a null model to determine if evolution has occurred in a population. Finally, you will perform computer simulations of population genetic processes during Week 6.

Lab Breakdown:

Week 4 – Examination of wild-type and mutant *Drosophila* phenotypes and sex identification. Begin blood typing exercise.

Week 5 – F₁ phenotyping, setting up F₁ crosses. Continue working on blood typing exercise.

Week 6 – Remove all adult F₁ flies from cross vials. Perform population genetic simulations.

Week 7 – Score phenotypes of all F₂ flies. Data analysis and review.

***Drosophila* Genetics**

The fruit fly, *Drosophila melanogaster*, is a ‘model organism’ that has been used for genetics research for many years. These flies have a relatively short generation time, are easy to cross, produce a large number of offspring, have relatively small genomes, and are harmless to humans, making them an ideal study system to understand patterns of inheritance in a controlled laboratory setting. We will be studying inheritance patterns of both wild-type and mutant phenotypes of flies over the course of several weeks. **In Week 5 you will cross F₁ flies exhibiting different phenotypic characteristics. Crosses may be either monohybrid or dihybrid. Your goal will be to determine which traits are dominant, recessive, autosomal, and sex-linked.** To help guide you through the activities it will be useful to think back to what you learned about Mendelian inheritance and Punnett squares.



Fly Lab Week 4 – Sexing Fruit Flies and Quantifying Phenotypes

For the first portion of the fruit fly exercise your goal is to become familiar with determining the sex of flies. Although this can be difficult at times, there are a few key criteria that can be used to help make the determination. Male flies generally have a darker posterior region versus females, whereas the female abdomen tends to show several distinct stripes (Fig. 1).

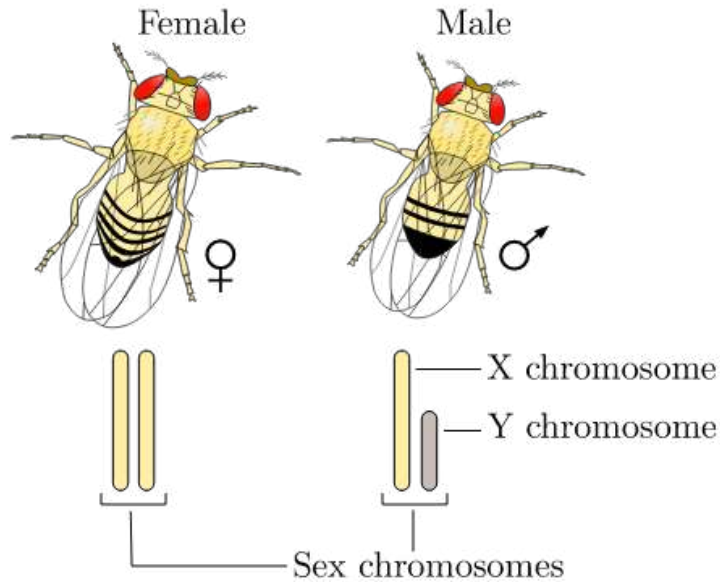


Fig. 1. A comparison of male and female fruit flies (*Drosophila melanogaster*). Males tend to show a dark, blunt posterior whereas females demonstrate light, pointed posteriors. *Image credit: CC BY-SA 4.0 YassineMrabet (author).*

In addition to determining the sex of your flies (with a stereomicroscope or magnifying glass), you are expected to become familiar with the different phenotypes present in mutant and wild-type varieties. **Pay particular attention to eye color and wing shape.** Your instructor will provide anesthetized flies and petri dishes for your analysis. **Fill in the table below with your observations and make sketches of the phenotypes and the sexes.** When finished, transfer your flies to the vial of alcohol.

Table 1. Phenotypes of wild-type and mutant parental strains.

Phenotype	Phenotype
Wild Type:	Mutant A:
Wild Type:	Mutant B:
Wild Type:	Mutant C:

Space for sketches:

Fly Lab Week 5 – Quantification of F₁ Phenotypes and Setting Crosses

This week you will be scoring F₁ phenotypes and setting up F₁ crosses to produce the F₂ generation. F₁ flies were already produced for you by crossing two pure strains together (from Table 1 above). Before you make your crosses, fill out the table below with your observations for F₁ flies from particular parental strains (P generation).

$$\frac{\text{Female Parent (P gen)}}{\text{Female Parent (P gen)}} \times \frac{\text{Male Parent (P gen)}}{\text{Male Parent (P gen)}}$$

Table 2. Sex-specific phenotypes of F₁ flies.

Date	Phenotype	# Females With Phenotype	# Males With Phenotype	Total Number of Flies

The next step is to setup your F₁ crosses. As a class, we need to make sure that all crosses are made in replicate to obtain a decent sample size.

- To setup your culture vials, add one cup of the Instant *Drosophila* Medium and 15 mL of water to the vial. When the solution solidifies, add 4-7 grains of yeast to the culture media.
- Add six (6) male/female pairs of anesthetized flies into your culture vial. **Be sure to label the vial with 'F₁ cross', the parental strains crossed, the date, and your names.**
- After all the F₁ flies are scored, the data are recorded, and the crosses set, dispose of any remaining anesthetized F₁ flies into the vial with alcohol.
- Based on the data in Tables 1 and 2, come up with a hypothesis for how your trait(s) are inherited. For example, how many genes are involved? Which alleles are likely dominant and recessive? Are any traits sex-linked?**
- Create two Punnett squares showing the results of both your P and F₁ crosses. You may use any letters you like for alleles. Results will vary depending on which strains your group selected for crosses.**
- What are the predicted phenotypic ratios of your F₂ generation?**



Fly Lab Week 6 – Removing F₁ Flies from Crossing Vials

The fly activities this week are relatively short. Thus, this is a good time to focus on the population genetics simulation lab and the associated questions and the end of the manual. At this point in the experiment you will need to remove the F₁ flies from your crossing vials. This is necessary to prevent mating between F₁ and F₂ flies, which will bias the results. Also, by Week 6 you should begin to see F₂ fly larvae emerging in your vials. **Place your culture vial under a dissecting microscope to get a better look at the morphological features of your larvae. If possible, make some sketches of your larvae using the space below. Make sure your flies don't escape!**

1. Obtain an empty culture vial and plug.
2. Tap your F₁ cross vial on your bench to knock the adult flies to the bottom. Remove plug and place mouth of empty vial on mouth of cross vial. Invert vials to transfer adult flies into the new vial. Quickly replace plug in both vials.
3. Anesthetize adult flies by dipping wand into Fly Nap anesthetic and inserting wand beside the plug.
4. When adequately anesthetized, place flies in alcohol vial.

Sketches of Larvae



Fly Lab Week 7 – Quantification of F₂ Phenotypes and Data Analysis

Your goal this week is to score the phenotypes of the F₂ flies that originated from your F₁ crosses. **Determine both the sex of each fly and the relevant phenotype(s).** Anesthetize the adults and transfer them to a petri dish for scoring. Use the dissecting microscope or magnifying glass to aid in data collection. Try to score as many flies as possible and fill out the table below.

Table 3. Sex and phenotypic data for F₂ offspring.

_____ X _____
 Female Parent (P gen) Male Parent (P gen)

Phenotype (and sex)	Total

Total scored _____

Next, fill out the corresponding tables with **class totals** for each cross. Results will vary depending on which crosses were performed.

Table 4. F₂ sex and phenotypic data for the entire class.

_____ X _____
 Female Parent (P gen) Male Parent (P gen)

Phenotype (and sex)	Total

Total scored _____



_____ X _____
Female Parent (P gen) Male Parent (P gen)

Phenotype (and sex)	Total

Total scored _____

_____ X _____
Female Parent (P gen) Male Parent (P gen)

Phenotype (and sex)	Total

Total scored _____

1. Calculate the observed phenotypic ratios for the class data for each cross.

_____ X _____
Female Parent (P gen) Male Parent (P gen)

F₂ phenotypic ratio:

_____ X _____
Female Parent (P gen) Male Parent (P gen)

F₂ phenotypic ratio:

_____ X _____
Female Parent (P gen) Male Parent (P gen)

F₂ phenotypic ratio:



2. Do the observed ratios correspond to the ratios you predicted? Discuss your reasoning.
3. Using the total observed class data for each cross, conduct chi-square analyses to determine if there is a statistically significant difference between observed and expected phenotypes. Remember, in a monohybrid cross we expect a 3:1 phenotypic ratio in the F₂ generation, whereas in a dihybrid cross we expect a 9:3:3:1 ratio. Recall the chi-square formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3}$$

The degrees of freedom (df) = # different F₂ phenotypes – 1. Thus, df = 1 for a monohybrid cross and df = 3 for a dihybrid cross. Use the chi-square table to determine if your calculated chi-square values are statistically significant.

Chi-Square Table

Table 5-2
Critical Values of the χ^2 Distribution

df \ p	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15

As an example, suppose you performed a monohybrid cross of wild-type and sepia-eyed flies and scored 100 F₂ flies. Out of the 100 flies, 65 had the wild-type phenotype and 35 had the mutant (sepia) phenotype. Are the results significantly different from expectations (i.e. 3:1 ratio)?

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(65 - 75)^2}{75} + \frac{(35 - 25)^2}{25} = 1.33 + 4 = 5.33$$



With 1 df and $P = 0.05$, a chi-square value of 5.33 is significantly higher than that expected by chance. Thus, we can conclude that the observed results are significantly different from expectations (i.e. we reject the null hypothesis of no statistically significant difference between observed and expected values).

Go ahead and perform your own series of calculations for each of the three crosses examined in class. Are any results consistent with expectations? Explain.



Human Blood Types

Human blood can be of four primary types: **Type A**, **Type B**, **Type AB**, and **Type O**. These types are determined based on a specific **antigen** present on an individual's red blood cells. For example, someone with Type A blood possesses Type A antigen on red blood cells whereas someone with Type B blood has Type B antigen on his/her red blood cells (Fig. 1). Individuals with Type AB blood show a phenotype expressing both Type A and Type B antigens. Finally, someone with Type O blood has neither Type A or Type B antigens on red blood cells. In addition, individuals with certain blood types will possess **antibodies** in their blood that will react with different blood types. This is why doctors make sure that they know your specific blood type before a transfusion is given! For example, a normal individual with Type A blood will possess antibodies *against* Type B blood. If this individual was given a transfusion of Type B blood, the anti-B antibodies in the patient's blood will bind the Type B antigen on the transfused blood, leading to **agglutination**, or the clumping of red blood cells. This can lead to reduced efficiency of oxygen transport in the body and death. Since individuals with AB blood have both Type A and Type B antigens, no antibodies to either are present in the blood. This is why blood type AB is considered the 'universal recipient.' They can receive blood from pretty much anyone without fear of a negative reaction. Conversely, people with Type O blood possess antibodies against both Type A and Type B. Thus, these individuals can only receive blood from others with Type O. However, since Type O blood cells do not contain any A or B antigen, any blood group can receive Type O blood. This is why blood type O is considered the 'universal donor.' Figure 2 provides a summary of donor and recipient status for different blood types. In addition to the ABO system, you will often see a + or – sign after the blood type. This designated whether the blood is positive or negative for a second antigen called a Rhesus or Rh-factor that is controlled by different genes.



	Group A	Group B	Group AB	Group O
Red blood cell type				
Antibodies in Plasma	 Anti-B	 Anti-A	None	 Anti-A and Anti-B
Antigens in Red Blood Cell	 A antigen	 B antigen	 A and B antigens	None

Fig. 1. Overview of the four major blood types found in humans. *Image credit: InvictaHOG (author).*

RED BLOOD CELL COMPATIBILITY TABLE								
Recipient	Donor							
	O-	O+	A-	A+	B-	B+	AB-	AB+
O-	✓	✗	✗	✗	✗	✗	✗	✗
O+	✓	✓	✗	✗	✗	✗	✗	✗
A-	✓	✗	✓	✗	✗	✗	✗	✗
A+	✓	✓	✓	✓	✗	✗	✗	✗
B-	✓	✗	✗	✗	✓	✗	✗	✗
B+	✓	✓	✗	✗	✓	✓	✗	✗
AB-	✓	✗	✓	✗	✓	✗	✓	✗
AB+	✓	✓	✓	✓	✓	✓	✓	✓

Fig. 2. Compatibility chart demonstrating which human blood types can serve as donors and recipients. *Image credit: CC BY 2.0 twintiger007 (author).*

Inheritance of human blood type deviates from simple Mendelian assumptions. The first deviation is that **three different alleles** control the blood phenotype in the ABO system. These alleles are commonly designated I^A , I^B , and i . The A and B alleles are **codominant**, meaning that both phenotypes are expressed in heterozygotes (Fig. 3). The i allele is recessive, meaning that individuals with Type O blood must have two copies to demonstrate the O phenotype.

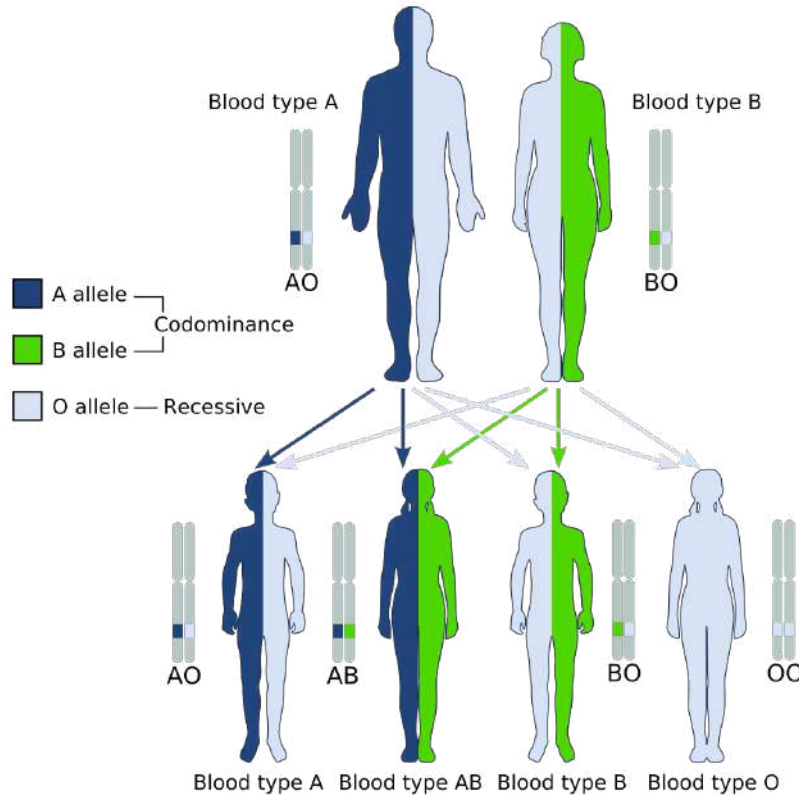


Fig. 3. Inheritance patterns within the human ABO system. Not that the A and B alleles are codominant and the O (or i) allele is recessive. *Image credit: YassineMrabet (author).*

Using the allelic nomenclature above, Table 1 shows all of the potential genotypes and phenotypes in this system.

Genotype	Phenotype
$I^A I^A$	A
$I^A i$	A
$I^B I^B$	B
$I^B i$	B
$I^A I^B$	AB
$i i$	O

Table 1. Possible genotypes and corresponding phenotypes in the human ABO blood group.

Use a Punnett square to conduct a cross between a heterozygous male with a $I^A I^B$ genotype with a heterozygous female with a $I^B i$ genotype. What are the resulting phenotypic ratios?

The **Hardy-Weinberg principle** is a foundational null model in population genetics that is used to determine if evolution has occurred in a population. If all of the Hardy-Weinberg assumptions are met, we can conclude that the population is not evolving. In reality, populations are always evolving due to one or multiple microevolutionary forces that change allele frequencies from generation to generation. **Remember, evolution is defined as changes in allele frequencies over multiple generations.** Microevolutionary forces include the following:

- a. **Mutation** – the spontaneous creation of a new allele in a population
- b. **Gene flow** – the transfer of alleles from one population to another
- c. **Natural selection** – the influence of the environment on relative fitness of individuals in a population. Some alleles in a population might be selected for, whereas others will be selected against.
- d. **Genetic drift** – changes in allele frequencies due to mating efficiency and randomness associated with allelic segregation during gametogenesis and meiosis (Fig. 4).
- e. **Nonrandom mating** – sometimes considered a microevolutionary force due to changes in genotype frequencies.

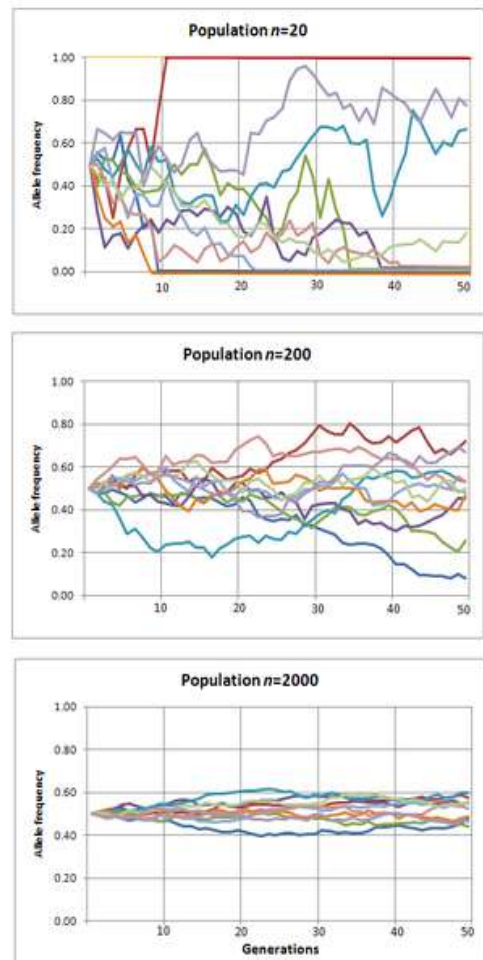


Fig. 4. Graphical depiction of genetic drift changing allele frequencies over multiple generations. Note the influence of population size on intensity of change. *Image credit: CC BY-SA 3.0 Professor marginalia (author).*

Let's explore the Hardy-Weinberg (HW) principle in more detail. Suppose we are studying a gene in a population that consists of two alleles, we will call *A* and *a*. I tell you that the *A* allele is dominant and the *a* allele is recessive. Using HW notation, we call the **A allele p and the a allele q**. You also know that the frequency of p is 0.73.

What is the frequency of q?

$$p + q = 1$$

$$q = 1 - p$$

$$q = 1 - 0.73 = 0.27$$

Next, we can predict the genotype frequencies in the population under the HW assumption. Because we are generally working with diploid individuals possessing two alleles per gene/locus, we can come up with a formula to estimate genotype frequencies from allele frequencies.

$$(p + q)(p + q) = p^2 + 2pq + q^2 = 1$$

where p^2 is the genotype frequency for the homozygote dominant genotype, $2pq$ is the genotype frequency of heterozygotes, and q^2 is the genotype frequency for the homozygote recessive.

Using the allele frequencies of A and a above, calculate the genotype frequencies expected under HW assumptions.

As stated above, the HW principle is used as a null model to determine if a population is evolving. In practice, researchers collect genotypic data from a population of interest and compare the real genotypic frequencies from the expected frequencies under HW. Significant deviations from HW assumptions suggest that one or more microevolutionary forces are changing allele and genotype frequencies over generations.

Let's try an example. Say you go out to a tropical island to study the population genetics of a species of lizard. For one of the genes you are examining you find only two alleles in the population, we will call T and t . You use molecular genetic tools to obtain the following genotypes for 100 lizards:

$$TT = 44$$

$$Tt = 36$$

$$tt = 20$$

Are these data consistent with HW expectations? Explain your reasoning and show your work. Raise your hand if you need assistance! **Hint: calculate p first!**



As with earlier problems, we can use a chi-square test to statistically test if the observed data differ from HW expectations. **For HW tests with two alleles we use 1 degree of freedom** because once we know p, q can be easily determined and vice versa. In the chi-square table the critical value at P = 0.05 is 3.84. Thus, calculated chi-square values > 3.84 suggest that the population is deviating significantly from HW expectations.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3}$$

Blood Type Activity (Weeks 4 – 5)

As stated above, we will be working with this lab for **two weeks** along with the fruit fly and simulation labs. Your goal is to use synthetic blood samples to determine blood genotypes of an island population of humans with unknown ancestry. Some island inhabitants suggest that they share ancestry with populations on a neighboring island, whereas others suggest that the majority of their genetic makeup comes from foreign populations. We will be performing multiple rounds of blood testing over the course of two weeks, keeping track of the total number of counts for each blood type. The blood dishes can be reused, but make sure you rinse them well before testing different samples. The mixing sticks may be thrown away after use. However, be sure to use the correct color stick depending upon the treatment (blue, yellow, clear).

1. Obtain a mixing tray and a blood sample from your instructor. Place a drop of blood into each of the three wells in your dish. Be sure to replace cap on sample.
2. Add drop of Anti A (blue vial) in well labeled 'A.'
3. Add drop of Anti B (yellow vial) in well labeled 'B.'
4. Add drop of Anti D (clear vial) in well labeled 'Rh.'
5. Use the appropriately colored mixing stick to mix the blood sample with the serum. **Make sure not to contaminate your samples!** Dispose of mixing sticks when finished.
6. Monitor wells to determine if agglutination occurs (this should hopefully be obvious). Fill in the table below with a 'yes' or 'no' depending on whether or not agglutination occurred. A positive reaction will indicate the blood type (e.g. a 'yes' for Anti-A only would signify an A- blood type). **Each student will test two samples each week.**

Week 4

	Sample #	Sample #
Anti-A		
Anti-B		
Anti-Rh		
Blood type		

Week 5

	Sample #	Sample #
Anti-A		
Anti-B		
Anti-Rh		
Blood type		

7. Each week we will make a class tally of the different blood types. Fill in each table below.

Week 4 (Island 1)

Blood Group	Totals for Group	Blood Group Frequency
A+		
A-		
B+		
B-		
AB+		
AB-		
O+		
O-		

Week 5 (Island 2)

Blood Group	Totals for Group	Blood Group Frequency
A+		
A-		
B+		
B-		
AB+		
AB-		
O+		
O-		

8. For each week, use the total class data to calculate the HW expected genotype frequencies for alleles controlling Rh factor. Assume that there are two alleles we will call *D* and *d*. Individuals with a *DD* or *Dd* genotype show the Rh+ phenotype and *dd* individuals are Rh-. **Hint: calculate q^2 and q first!**



9. The tables below shows the blood group frequencies of inhabitants of two different islands neighboring the islands that you have already scored above. What can you conclude based on these blood group frequencies and the genetic makeup of all islands?

Island 3

Blood Group	Blood Group Frequency
A+	0.31
A-	0.05
B+	0.01
B-	0.01
AB+	0.01
AB-	0.01
O+	0.43
O-	0.17

Island 4

Blood Group	Blood Group Frequency
A+	0.07
A-	0.22
B+	0.1
B-	0.1
AB+	0.08
AB-	0.2
O+	0.08
O-	0.15

10. For the two additional island populations in Question 9, calculate the expected Rh factor genotype frequencies under HW equilibrium. How do the results compare to the two populations you blood typed? Explain your reasoning.

11. So far, we have only examined HW principles in genes consisting of two alleles. We can extend the HW equations to more than two alleles, although the math can get confusing quickly. However, it isn't much more difficult to calculate expected frequencies with three alleles. Our blood type activity provides an excellent example as the ABO type is controlled by three alleles: I^A , I^B , and i .

With three alleles:

$$p + q + r = 1$$

and thus,

$$(p + q + r)(p + q + r) = p^2 + 2pq + 2pr + q^2 + 2qr + r^2 = 1$$

where

p^2 = expected frequency of $I^A I^A$ genotype

q^2 = expected frequency of $I^B I^B$ genotype

r^2 = expected frequency of ii genotype

$2pq$ = expected frequency of $I^A I^B$ heterozygotes

$2pr$ = expected frequency of $I^A i$ heterozygotes

$2qr$ = expected frequency of $I^B i$ heterozygotes

Suppose you go out and obtain blood genotypes from 1000 people throughout NYC and obtain the following results:

Genotype	Count	Genotype frequency
$I^A I^A$	210	0.21
$I^B I^B$	75	0.075
ii	330	0.33
$I^A I^B$	143	0.143
$I^A i$	86	0.086
$I^B i$	156	0.156

Perform a chi-square analysis to determine if these results are consistent with HW expectations. In this case, because we are dealing with three alleles we will need to use 2 degrees of freedom. **At $P = 0.05$ the critical chi-square value is 5.991.** Remember, to solve HW problems you must first calculate allele frequencies (p , q and r in this case) and use the allele frequencies to calculate expected genotype frequencies under HW. **Show all work!**

12. Cystic fibrosis is an autosomal recessive disease caused by a mutation in a single gene with two alleles. Suppose the frequency of cystic fibrosis in a population is 3%. Assuming the gene is in HW equilibrium, calculate the allele and genotype frequencies. If the population consists of 100 individuals, how many are carriers (but unaffected) for the cystic fibrosis allele?

Review Questions

1. What is meant by simple Mendelian inheritance? Discuss some additional patterns of inheritance that would lead to phenotypic ratios that deviate from simple Mendelian patterns.
2. Discuss patterns of inheritance in the human ABO blood system. What would be the expected phenotypic ratios of offspring between someone with Type-O blood and someone who was heterozygous for Type-A blood?
3. Why is Hardy-Weinberg equilibrium such a useful concept in population genetics and evolution? Which processes may lead to deviations from Hardy-Weinberg expectations? Which of these processes are considered adaptive?
4. Discuss why/how we are able to determine the expected genotype frequencies (under Hardy-Weinberg) at the Rh gene in our blood typing experiment.



5. What phenotypic traits did we examine with *Drosophila*? Which crosses were monohybrid and which were dihybrid? What were we testing with these experiments? What was your null hypothesis?

6. Assume that a disease is caused by a mutation at a single gene and is inherited in a recessive manner. Only two alleles are present, the wild-type and the mutant. In a population of 567 people, 76 are affected with the disease. Assuming the gene is in Hardy-Weinberg equilibrium, calculate the expected allele and genotype frequencies. What proportion of **unaffected** individuals are heterozygotes?

7. Why are *Drosophila* commonly used in genetics research?

8. Discuss why the determination of human blood group frequencies may help determine evolutionary relatedness of populations. In your response, state how different microevolutionary processes interact to shape population allele frequencies. Can you think of any potential limitations of inferring evolutionary history solely from data at the ABO gene? How might random sampling error come into play?

9. Do any of the *Drosophila* traits studied suggest sex-linked inheritance? Assume that we are studying eye color in *Drosophila*, which is assumed to be controlled by a gene on the X-chromosome. The wild-type eye color is red, whereas the mutant is white. What are the F₂ phenotypic ratios in a cross between a red-eyed male and a heterozygous carrier red-eyed female? What proportion of males would exhibit the mutant phenotype?
10. Suppose you crossed two flies, one of which was homozygous for red eyes and one of which was homozygous for sepia eyes. You then mate the F₁ generation to each other to obtain the F₂ generation. The number of F₂ flies with red eyes is 357 and the number with sepia eyes is 112. Are these data consistent with simple Mendelian expectations? Explain your reasoning and show all your work.

Simulation of Population Genetic Processes (Week 6)

Background:

Once considered a strictly theoretical discipline, technological advances have now made it feasible to assay genetic variation in natural populations. Quantification of population genetic structure has importance within many biological disciplines including human genetics, medicine, evolution, ecology, and conservation. In general, the field of population genetics is focused on determining the processes responsible for shaping spatial patterns of allelic variation. Recall that evolution, specifically **microevolution**, can be defined as changes in population allele frequencies over multiple generations. Four primary mechanisms are responsible for controlling allele frequencies within populations: **mutation**, **gene flow**, **natural selection**, and **genetic drift**. Non-random mating (including sexual selection) does not directly affect allele frequencies, but can influence genotype frequencies within populations. Brief definitions of each microevolutionary process are provided below.

- f. **Mutation** – the spontaneous creation of a new allele in a population
- g. **Gene flow** – the transfer of alleles from one population to another
- h. **Natural selection** – the influence of the environment on relative fitness of individuals in a population. Some alleles in a population might be selected for, whereas others will be selected against.
- i. **Genetic drift** – changes in allele frequencies due to mating efficiency and randomness associated with allelic segregation during gametogenesis and meiosis.
- j. **Nonrandom mating** – sometimes considered a microevolutionary force due to changes in genotype frequencies.

It is important to make the distinction between what are called **neutral microevolutionary processes versus adaptive microevolutionary processes**. Natural selection is considered an adaptive process because it alters allele frequencies by ‘selecting’ for certain alleles that confer a fitness advantage to individuals possessing them. For example, beak morphology in Darwin’s finches differs to best utilize different food sources that naturally occur on different islands of the Galapagos (Fig. 1). Conversely, genetic drift, gene flow, and mutation are often considered neutral processes because they change allele frequencies through non-adaptive mechanisms.

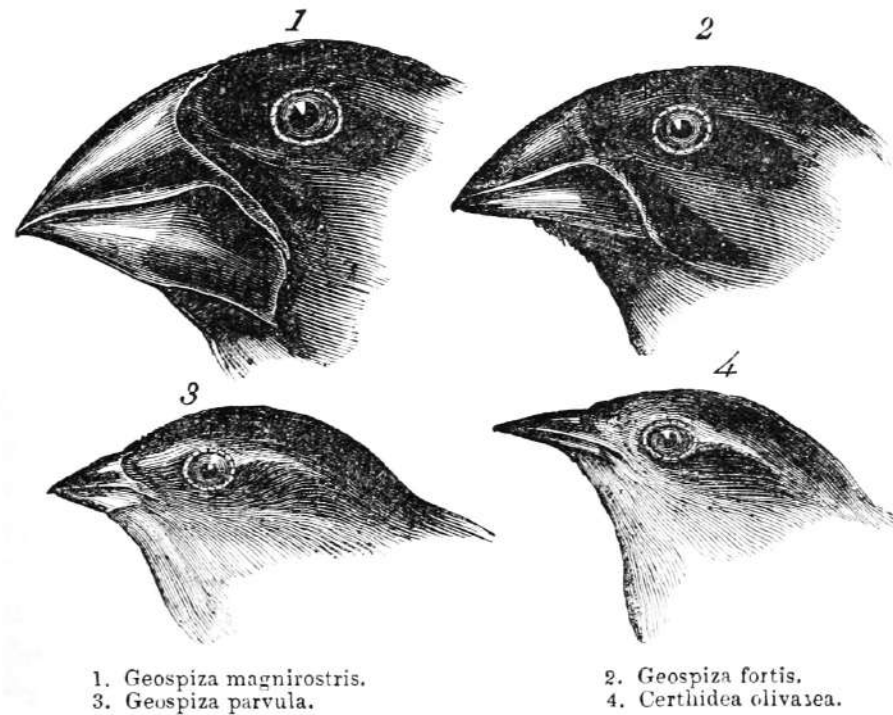


Fig. 1. Example of natural selection using Darwin's finches. Different beak morphologies evolved in different species inhabiting different islands to best utilize natural food sources. In other words, different habitats 'selected' for birds with certain phenotypic characteristics. *Image credit: John Gould.*

In real populations it is important to note that many times multiple microevolutionary processes act in synergy to control allele frequencies. For example, a new random mutation may originate in a population of lizards leading to a new allele. This new mutation may confer some selective advantage over other alleles already present in the population. Because individuals possessing this allele are more fit, natural selection will increase the frequency of the new allele (Fig. 2).

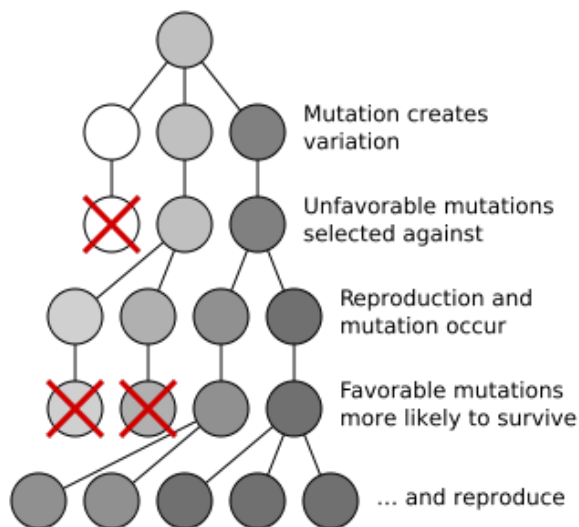
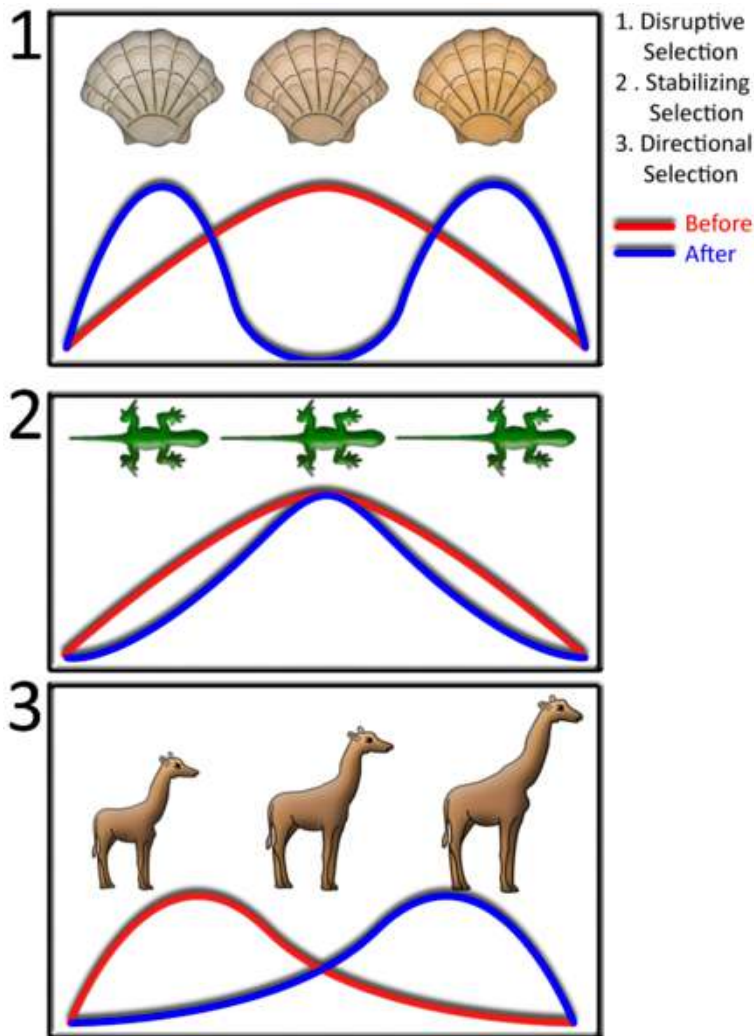


Fig. 2. Example of how mutation and natural selection may interact to shape allele frequencies in populations. *Image credit: CC BY-SA 3.0 Elembis (author).*

Evolutionary biologists can further classify different types of natural selection based on how traits and allele frequencies shift. Figure 3 highlights some of the more common forms of selection. In **disruptive selection**, phenotypes at opposite ends of a continuum are selected for. This is often a result of heterogeneous habitat where one form has higher fitness in one ecological extreme versus the other. In **stabilizing selection**, intermediate phenotypes have the highest fitness. Stabilizing selection tends to decrease the phenotypic variance in



populations. **Directional selection** favors the evolution of one specific extreme phenotype. For example, larger giraffes may exhibit higher fitness because they are better able to access food sources and reproduce.

Random mutations occur so infrequently that population geneticists and evolutionary biologists do not consider it to be a major driving force shaping allele frequencies in populations. However, mutation **IS** an important source of **new** genetic variation. Researchers have argued for decades on the relative importance of neutral versus adaptive processes in evolution. The Darwinian view posits that natural selection (i.e. adaptive processes) is the primary mechanism to explain allele frequencies and the evolution of biodiversity. This was a widely accepted hypothesis since Charles Darwin came up with his theory of evolution by natural selection. Decades later, a biologist and researcher named Motoo Kimura suggested that most evolutionary changes are neutral and do not confer a fitness advantage or disadvantage to organisms. Kimura's **neutral theory of molecular evolution** placed a much stronger emphasis on genetic drift to explain the evolution of life on Earth. So what is the correct answer? Are neutral or adaptive processes more important? For the majority of organisms the most likely answer is that both drift and selection are important in shaping allele frequencies in

Fig. 3. Graphical depictions of different forms of natural selection. 1 = disruptive selection; 2 = stabilizing selection; 3 = directional selection. The original phenotypic distribution in populations is illustrated by the red curve, whereas the blue curve shows the phenotypic distribution following each type of selection. *Image credit: CC BY-SA 3.0 Azcolvin429 (author).*

populations, and both processes were/are vital in shaping the evolution of biodiversity.

Because genetic drift is such an important process in evolution and population genetics it is worthwhile to spend a bit of time to make sure that it is understood correctly. Remember that genetic drift changes allele frequencies by random processes. For example, who mates with whom and which alleles are passed to which gametes during meiosis. In general, genetic drift has a more profound influence in **smaller** populations (Fig. 4).



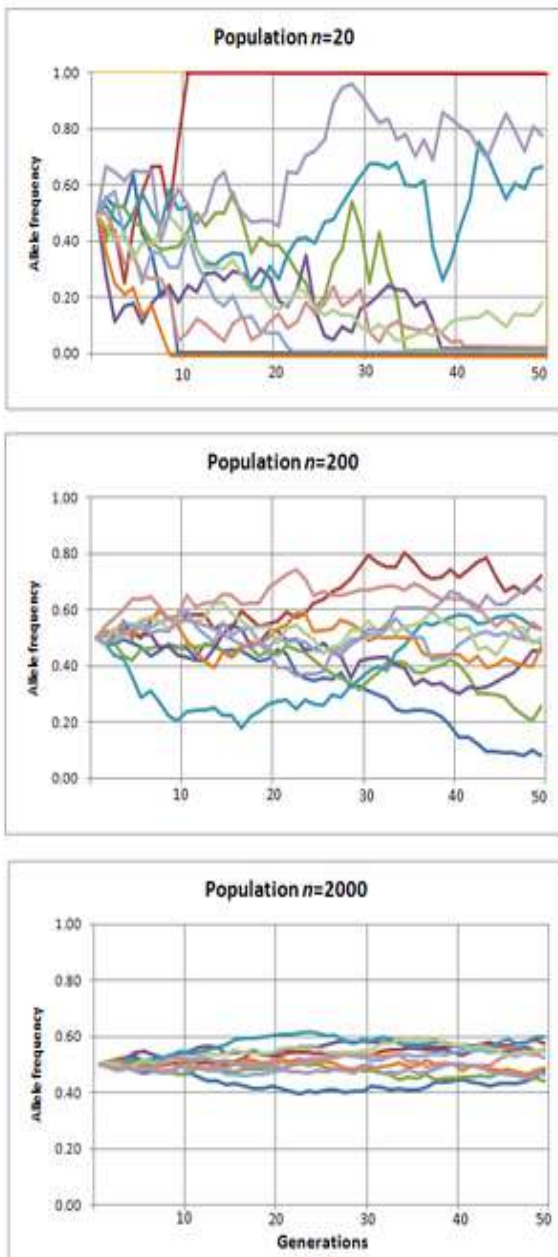


Fig. 4. The interplay between population size and genetic drift. Drift has a more profound affect in smaller populations and may lead alleles to become lost or fixed. *Image credit: CC BY-SA 3.0 Professor marginalia (author).*

Two additional concepts related to genetic drift are **bottlenecks** and **founder events**. In bottleneck events, some external force can substantially reduce the number of individuals (and alleles) in a population. Because genetic drift has a stronger influence in smaller populations, allele frequencies will drift to a greater degree following a bottleneck, leading some alleles to become fixed and others to be lost (Fig. 5). A similar process occurs during founder events, which can be defined as a small subset of individuals in a population colonizing a new location. A good real-world example is the colonization of Madagascar from mainland Africa. Research has shown that the majority of organisms colonized Madagascar by overseas dispersal (versus the continental separation of Africa and Madagascar ~165 Mya). Founder events can lead to what is called **founder event speciation, or peripatric speciation**. How do these founder events lead to the creation of a new species? Because founder events by definition involve a relative small number of individuals from a parental population, genetic drift will have a pronounced influence on allele trajectories once these individuals colonize a new area (Fig. 6). Similar to population bottlenecks, founder events can lead to the rapid fixation or loss of alleles, in turn increasing the genetic divergence between the parental and founder populations. Other types of speciation include **allopatric speciation**, which occurs when a geographic barrier (e.g. a mountain range) subdivides an ancestral population, **parapatric speciation**, which occurs along environmental clines, and **sympatric speciation**, which occurs without an obvious barrier to gene flow (Fig. 7).

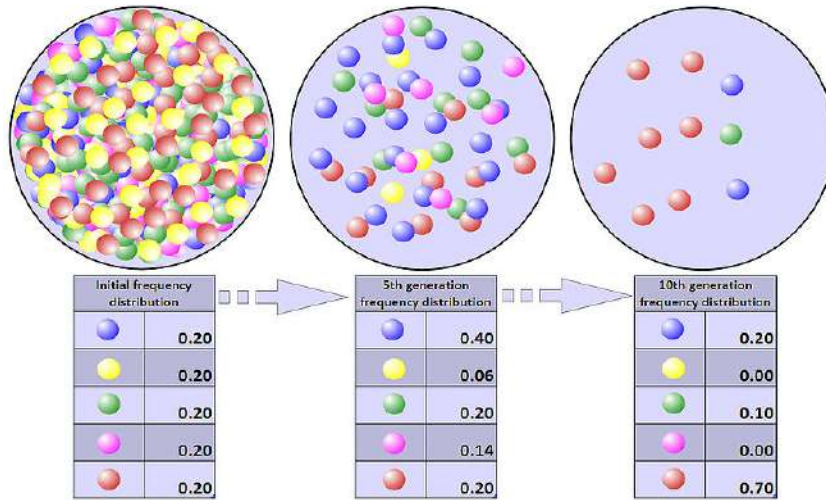


Fig. 5. Hypothetical example of how population bottlenecks may lead to the loss or fixation of alleles due to genetic drift. The original population of the left contained 500 alleles coded by color. Allele frequencies changed substantially following two bottleneck events during the 5th generation (center) and 10th generation (right). Note that by the 10th generation the yellow and pink alleles have been completely lost from the population. *Image credit: CC BY-SA 3.0 Professor marginalia (author).*

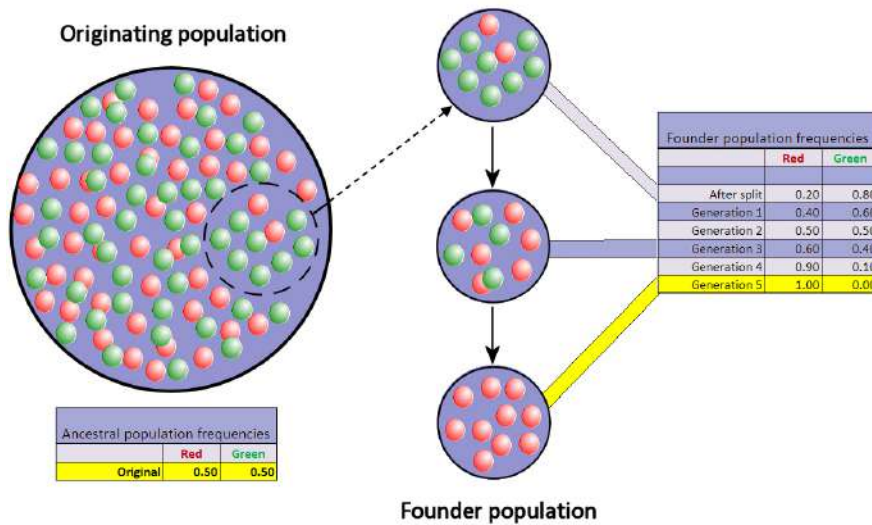


Fig. 6. Founder events occur when a small subset of individuals breaks off from a parental population to colonize a new area. As the founding population is small, genetic drift will exert a strong force on allele frequencies, causing some alleles to be lost and others to become fixed. In this example, the green allele is completely removed from the founder population after five generations. *Image credit: CC BY-SA 3.0 Professor marginalia (author).*

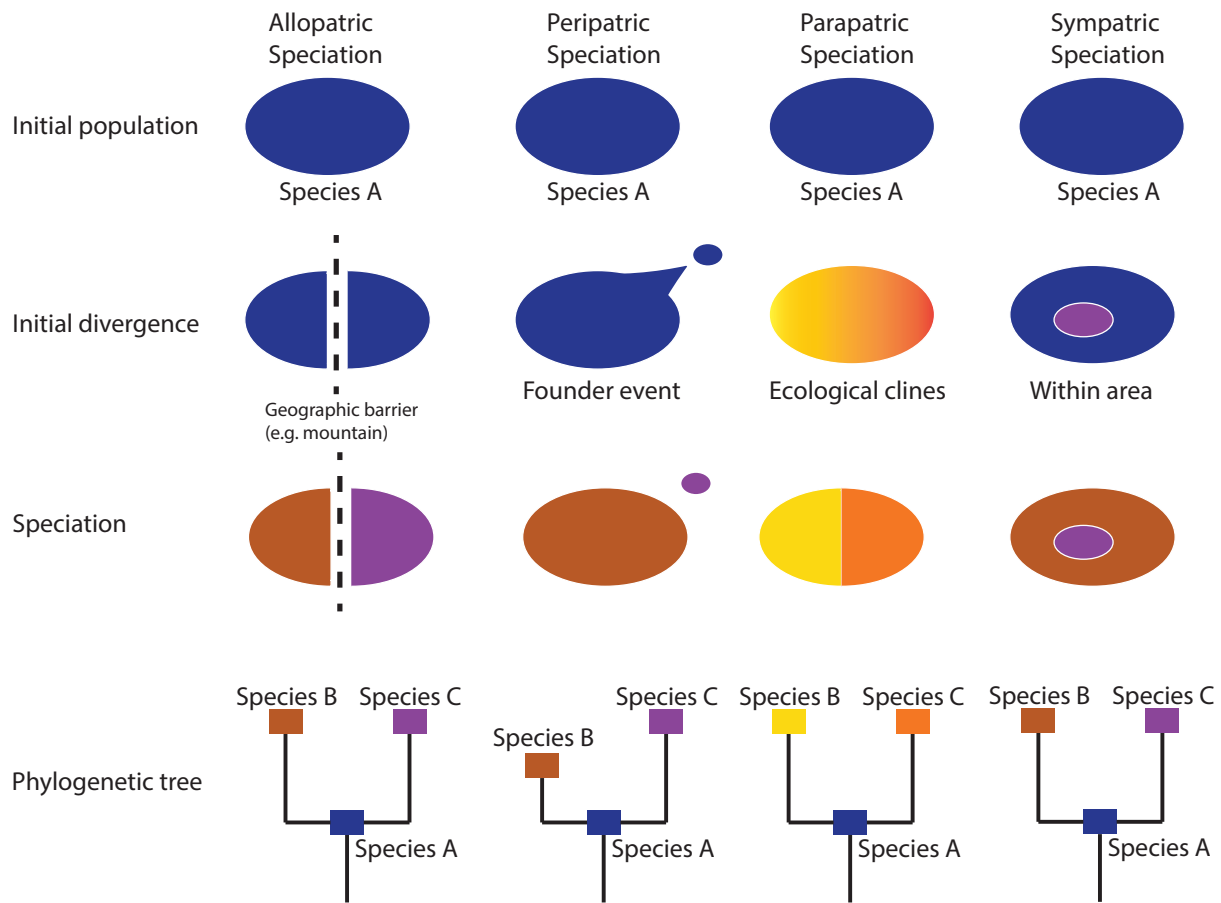


Fig. 7. Simplistic illustration of the four primary mechanisms of speciation. Allopatric and peripatric speciation require some type of physical barrier that leads to reproductive isolation of populations. Although similar, allopatric and peripatric divergence are distinguished based on population sizes. In peripatric speciation, a small group of individuals from the parent population colonizes a new geographic area. This population then experiences rapid genetic changes due to drift and selection. In parapatric speciation, a population is distributed along an environmental gradient. Divergent natural selection at opposite ends of the gradient results in the evolution of different phenotypic characteristics. Positive assortative mating (mating among similar individuals) eventually leads to reproductive isolation and speciation. Sympatric speciation results without physical or ecological barriers. Examples include the evolution of polyploidy in plants, and different host plant use by insects.

Activity:



A. Simulating the influence of genetic drift

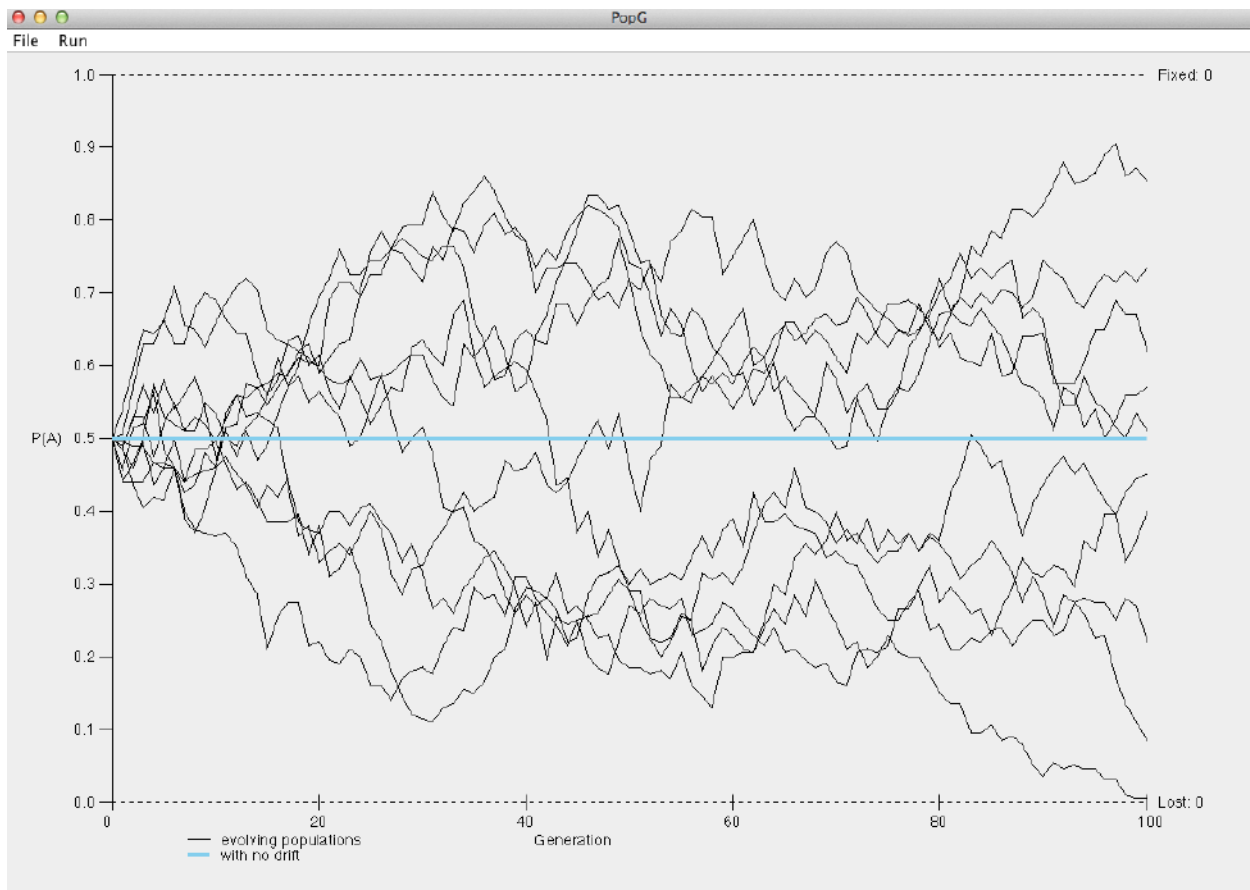
In this lab we will be working with a population genetic simulator called PopG (<http://evolution.gs.washington.edu/popgen/>). This is a relatively simple, but useful package that can be used to visualize how different microevolutionary processes shape allele frequencies over multiple generations. For simplicity, we will initially examine different processes individually. Go ahead and open the software where you should see a window for specifying certain parameters (see below).

Parameter	Value
Population size:	100
Fitness of genotype AA:	1.0
Fitness of genotype Aa:	1.0
Fitness of genotype aa:	1.0
Mutation from A to a:	0.0
Mutation from a to A:	0.0
Migration rate between populations:	0.0
Initial frequency of allele A:	0.5
Generations to run:	100
Populations evolving simultaneously:	10
Random number seed:	(Autogenerate)

Take a few minutes to familiarize yourself with some of the options. You will notice that there are options to control each of the microevolutionary processes: genetic drift, natural selection, mutation, and gene flow.

Why is genetic drift controlled by the Population Size parameter?

Hit OK to run the default settings. Hopefully you obtain a plot like the one below.



Look at the two axes. The x-axis plots generation number whereas the y-axis shown the frequency of the *A* allele. Each line in the figure represents a different population (you simulated 10).

How would you explain what is happening in these plots?

Go back to the settings window and change the number of generations from 100 to 1000.

What happened to the *A* allele in several of the populations?

Next, change the population size from 100 to 10000 and run the simulation.

How does the population size parameter influence the frequency of the A allele in each population? Does the A allele become fixed or lost in any population? How does this compare to the simulation using a population size of 100?

B. Simulating the influence of natural selection

Return to the settings window and click the 'Defaults' button to restore the default settings. We will now test how **natural selection** might influence the frequency of the A allele. To minimize the influence of drift in these simulations specify a large population size of 10000. When examining natural selection we usually specify the **relative fitness (w)** of the different genotypes. By default, a value of 1.0 is assigned to the genotype with the highest fitness, and the remaining genotypes are assigned values relative to the value of 1.0. Let's start by assigning the following relative fitness values:

$$w_{AA} = 1.0$$

$$w_{Aa} = 0.7$$

$$w_{aa} = 0.3$$

We can also introduce another parameter called the **selection coefficient (s)**. The selection coefficient is equal to $1 - w$. Using the relative fitness values above:

$$s_{AA} = 1 - 1.0 = 0$$

$$s_{Aa} = 1 - 0.7 = 0.3$$

$$s_{aa} = 1 - 0.3 = 0.7$$

Run the simulation for 10 generations using the relative fitness values above. **What happens to the frequency of the A allele?**

Next, run the simulation for 100 generations. You should notice that the A allele eventually becomes fixed in each population. **At approximately which generation does the A allele become fixed?**



In the previous example the homozygous dominant genotype exhibited the highest fitness. What would the simulation look like in a case of overdominance (i.e. heterozygote advantage)? Let's change the relative fitness values to the following:

$$w_{AA} = 0.6$$

$$w_{Aa} = 1.0$$

$$w_{aa} = 0.4$$

The selection coefficients in this case would be

$$s_{AA} = 1 - 0.6 = 0.4$$

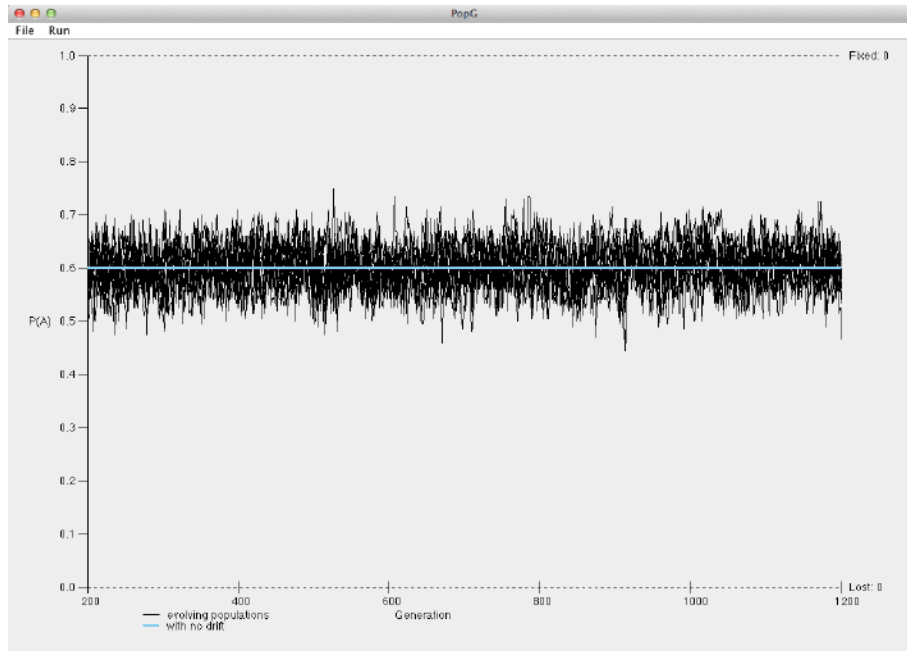
$$s_{Aa} = 1 - 1.0 = 0$$

$$s_{aa} = 1 - 0.4 = 0.6$$

Create a new simulation by first resetting the default parameters. Enter the relative fitness values above to model our example of overdominance. Again, play with the population size to determine the influence of drift. Leave all other settings as the defaults.

What do you notice about the frequency of the A allele over the course of 100 generations? What about for 1000 generations? Hopefully your plot looks something like the plot below.





How do your simulations of overdominance compare to your simulations assuming that the homozygous dominant genotype had the highest fitness?

C. Simulating the influence of gene flow

Gene flow is an extremely important microevolutionary process that can introduce new alleles into populations. For example, a beneficial allele lost due to genetic drift may be reintroduced to a population through gene flow. Gene flow can be defined as the successful migration of individuals between populations followed by the transfer of genetic information (i.e. breeding). Studying and quantifying rates and patterns of gene flow has wide-ranging implications for many fields including disease surveillance, human migratory patterns, conservation biology, and systematics. We can use the following equation to determine the change in allele frequency due to migration:

$$\Delta p = m(p_D - p_R)$$

where

Δp = the change in allele frequency in the recipient population

m = proportion of migrants within the recipient population

p_D = allele frequency of donor population



p_R = allele frequency of recipient population

When simulating gene flow it is often interesting to model the **interaction between migration and genetic drift**. Recall that genetic drift tends to promote the loss or fixation of alleles (leading to reductions in genetic diversity), particularly within small populations. This may lead to large genetic differences between populations, ultimately leading to **speciation**. Conversely, gene flow can help homogenize populations and maintain a stable number of alleles.

Start a new simulation by restoring the default settings. Click OK to run the simulation for 100 generations. You should see that the frequency of the A allele fluctuates considerably due to genetic drift alone (all other microevolutionary processes were controlled for). What happens if you introduce some migration between each of your 10 populations? Open up the settings window and enter a value of **0.3 for migration rate between populations**.

Discuss how the introduction of gene flow influenced allele frequencies of the A allele.

Do you think that gene flow would have a larger or smaller affect on allele frequencies in large populations? Explain your reasoning. It may be useful to perform additional simulations if needed.

D. Simulating the influence of mutation

Mutation is the ultimate source of genetic variation in populations. Mutation can create new alleles in populations, which can then be acted upon by other microevolutionary processes such as drift and selection. It's important to note that although mutation can create new alleles, it does not substantially alter allele frequencies because mutation is a relatively infrequent event. Mutation rate is commonly represented by the Greek symbol μ and can be expressed in units such as generations or time (e.g. per million years). For example, the typical mutation rate of vertebrate mitochondrial DNA (mtDNA) ranges between about 0.8 – 2.0 **substitutions per site per million years**. Site refers to a specific base position in an alignment of homologous DNA sequences. Mutation rates of animal nuclear DNA (nDNA) are often an order of magnitude slower. Interestingly, plant mtDNA evolves at a much slower rate than plant nDNA. Thus, different genomes have the ability to provide different amounts of evolutionary information, and target genes can be tailored to one's specific research objectives. We can use the equation below to calculate the frequency of the A allele after a specific number of generations, assuming we know the starting allele frequency and the mutation rate.

$$(1 - \mu)^t = \frac{p_t}{p_0}$$



where

μ = mutation rate from the *A* allele to the *a* allele

t = time (e.g. number of generations)

p_t = frequency of the *A* allele after t generations

p_0 = initial frequency of the *A* allele

We can rearrange the problem to solve for p_t :

$$p_t = p_0 * (1 - \mu)^t$$

Calculate the frequency of the *A* allele after 500 generations, assuming a mutation rate of 0.0001 per generation and an initial frequency of 0.5.

Let's perform some simulations to observe how mutation alters allele frequencies. Setup a new run by first restoring the defaults. To minimize the influence of drift, set the population size to 10000. Enter 0.0001 for the mutation rate from *A* to *a* and run the simulation for 500 generations.

What do the results suggest? Would you say that the frequency of the *A* allele is changing dramatically? Explain your reasoning.

Next, go back to the settings window and enter a rate of 0.001 for the mutation rate from *a* to *A*. and rerun the simulation.

How do these results differ from the previous simulation that assumed no mutation from *a* to *A*?

Finally, let's try to model a situation where mutation creates a new beneficial allele that is acted upon by natural selection. Theory predicts that the beneficial allele should become fixed in a population if the population size is large enough to minimize the affect of drift. In this example, we will assume that the population was originally monomorphic for the *a* allele, and mutation introduces a beneficial *A* allele. We will also assume that the mutation rate from *A* to *a* is the



same as the mutation rate from a to A (0.00001). We will use the following relative fitness values:

$$w_{AA} = 1.0$$

$$w_{Aa} = 0.7$$

$$w_{aa} = 0.3$$

Your settings should look like the diagram below.

Setting	Value
Population size:	10000
Fitness of genotype AA:	1.0
Fitness of genotype Aa:	0.7
Fitness of genotype aa:	0.3
Mutation from A to a:	1.0E-5
Mutation from a to A:	1.0E-5
Migration rate between populations:	0.0
Initial frequency of allele A:	0.0
Generations to run:	100
Populations evolving simultaneously:	10
Random number seed:	(Autogenerate)

Buttons: Defaults, Cancel, OK

Provide a detailed description of what is happening in your model.

If time permits, perform additional simulations of your choosing, mixing and matching different processes to see how allele frequencies are affected. Upon successful completion of the lab you should be comfortable with how the different microevolutionary processes interact to shape allele frequencies in populations.



DNA Fingerprinting

Background:

Throughout the first half of the semester we focused our efforts on classic genetic experiments including crosses. We will now turn our attention to different types of **molecular genetic techniques** that can be used to assess levels of genetic variation. The number of molecular genetic techniques available to researchers is quite large, and it is imperative that the correct method is used to address the question of interest.

DNA fingerprinting is a widely used approach to identify organisms and determine evolutionary relatedness. As an example, DNA fingerprinting is commonly used in forensics to identify criminals from a set of possible suspects. In short, residual traces of DNA are collected from a crime scene along with DNA samples from multiple suspects. DNA fingerprints are then constructed from the crime scene DNA and compared to all other DNA samples (Fig. 1). Matches in banding patterns between the crime scene sample and one of the suspects would implicate that particular suspect in the crime.

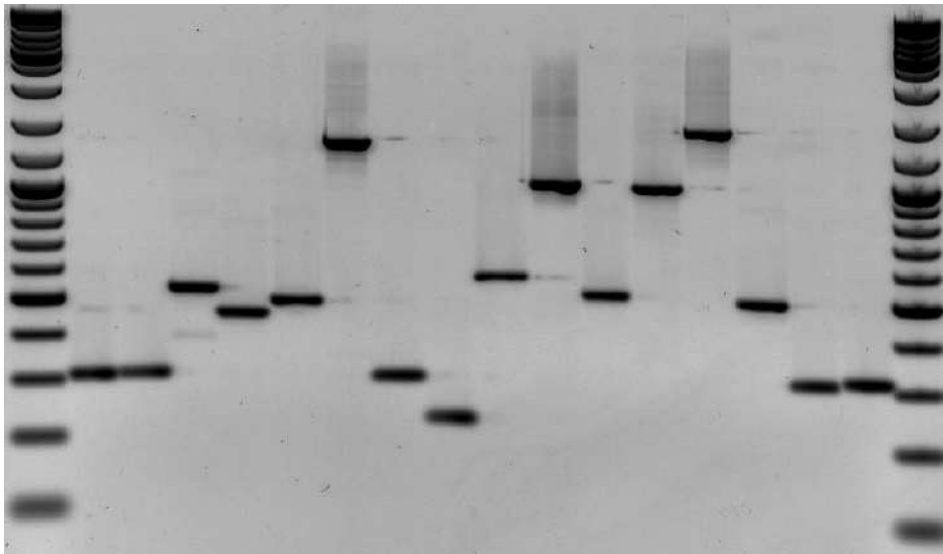


Fig. 1. Example of DNA fingerprinting. Each band on an agarose gel indicates a specific fragment of DNA. Each vertical column constitutes a different sample. Wells on the far left and far right contain a DNA ladder illustrating bands of known size. DNA ladders are used to help determine the size (in base pairs) of sample DNA fragments. *Image credit: CC BY-SA 3.0 Rkalendar (author).*

The technique can also be used to indicate paternity in a set of potential fathers. Figure 2 depicts the results of a DNA fingerprinting experiment focused on determining which of three men was the father of a child. 'M' represents the DNA fingerprint of the mother, 'Ch' is the fingerprint of the child, and samples 1, 2 and 3 are the fingerprints from the three potential fathers. Remember that in diploid organisms (including humans), there are two alleles per locus. **Looking at the banding pattern below, which male is the most likely father?**

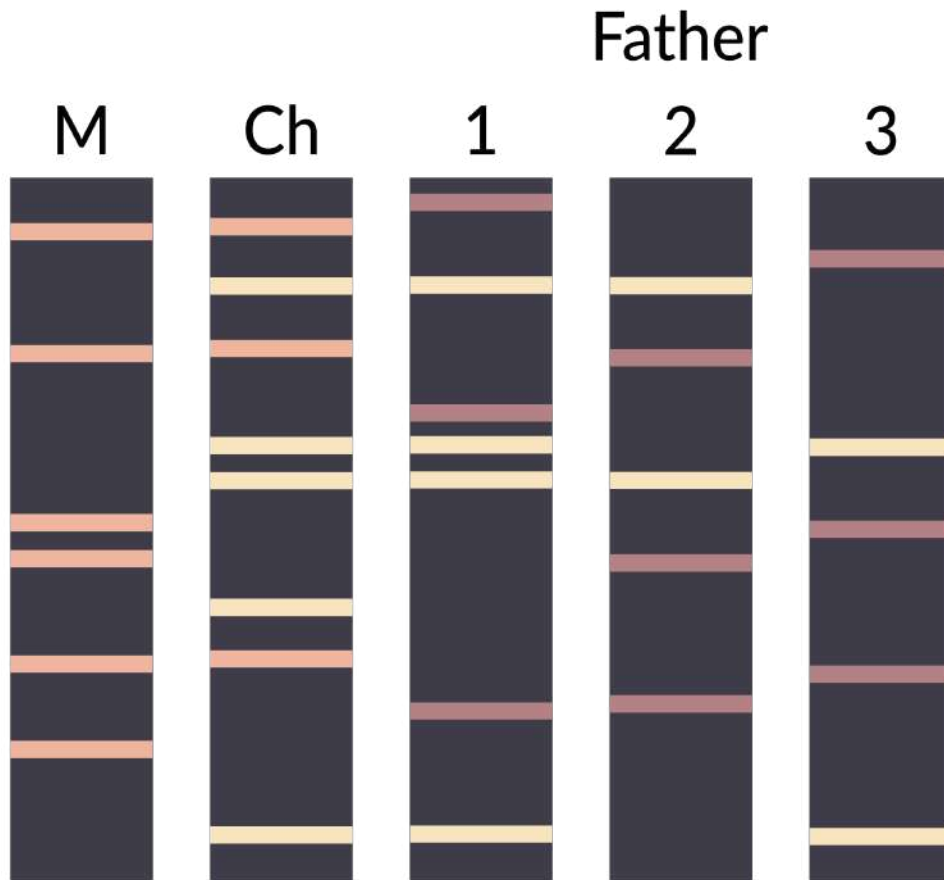


Fig. 2. Simplified example of how DNA fingerprinting techniques can be used to determine paternity among multiple candidate fathers. M = mother; Ch = child; 1,2 and 3 = potential fathers. *Image credit: CC BY-SA 3.0 Helixitta (author).*

As alluded to above, multiple molecular genetic techniques may be used to construct a DNA fingerprint to use for identification or relatedness purposes. One historically common method is termed **restriction fragment length polymorphisms (RFLPs)**. In this method, one or multiple **restriction enzymes** are used to cut the DNA at specific regions of the genome. Most restriction enzymes are derived from different species of bacteria and cleave DNA at specific recognition sites. Figure 3 illustrates the process of DNA digestion by the enzyme HindIII. This particular enzyme recognizes and cuts the following sequence:



Notice that following digestion, the cut DNA protrudes as overhangs that are called sticky ends. This contrasts with restriction enzymes that result in blunt-ended DNA fragments lacking an overhang. Once DNA is cut with a restriction enzyme, the fragmented DNA can be reattached through an enzyme called DNA ligase. Alternatively, a **plasmid** could also be cut with the same restriction enzyme, and the DNA fragment of interest may be inserted into the plasmid. The recombinant plasmid can then be introduced into a bacterial cell (e.g. *E. coli*) and continuously replicated during binary fission. This is part of the technique termed **molecular cloning**.

The first step in a typical fingerprinting analysis is to digest different DNA samples with a restriction enzyme. Next, the resulting fragments are separated using a technique called **horizontal agarose gel electrophoresis**. Digested DNA samples are first loaded into separate wells of an agarose gel. The DNA is generally mixed with a loading dye to assist in loading the samples into the gel. A **stain** is needed to actually visualize DNA banding patterns on the gel (i.e. DNA by itself is colorless). Along with the digested DNA samples, we want to load a **DNA ladder** (usually in the first well) to give us a broad idea of fragment sizes (Fig. 4).

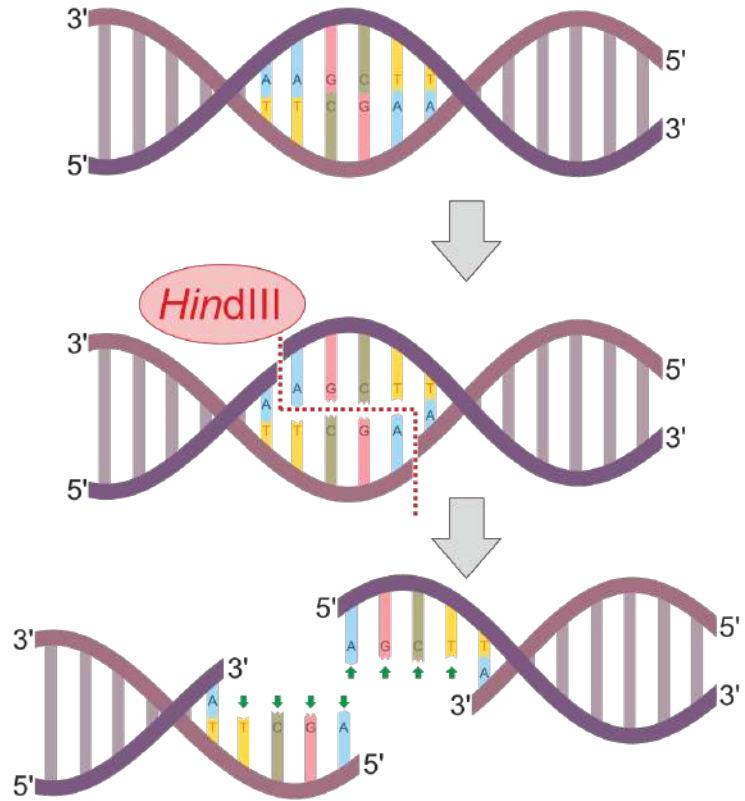


Fig. 3. Graphical depiction of DNA digestion using the restriction enzyme HindIII. Note the production of overhangs following digestion. *Image credit: CC BY-SA 4.0 Helixitta (author).*

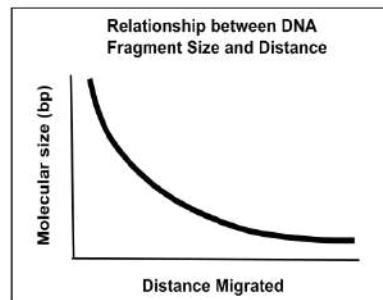
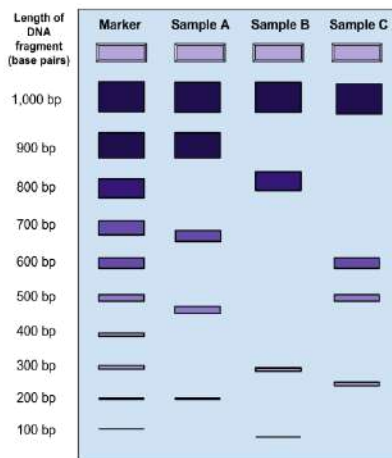


Fig. 4. DNA electrophoresis example. Gel illustrates three different banding patterns from three different samples. The marker is included to help determine the size (bp) of each fragment. Figure to the right shows the inverse relationship between distance migrated in the gel versus fragment size. *Image credit: CC BY-SA 4.0 Mckenzielower (author).*

Once all the samples are loaded in the gel inside of the gel rig, we apply an electrical current to 'electrophorese' the DNA through the gel. DNA is a negatively charged molecule and will migrate towards a positive electric field (**the red side of the gel apparatus**). DNA migrates at a rate inversely proportional to the size of the fragment (i.e. shorter fragments migrate at a faster rate). As you can see from Fig. 4, fragments that are 1,000 bp in size do not migrate far from the well where the sample was loaded. Depending upon the specific application, you'll want to run the gel for ~100 V for 30 min or so. As a general rule of thumb, I tend to stop the current once the DNA has migrated about halfway down the gel.

Note that in a more typical RFLP-based DNA fingerprinting experiment, a subsequent step called **Southern blotting** is needed in order to visualize and compare fragments. The reason is that when human DNA (or DNA from any other eukaryote with a large genome) is cut with a restriction enzyme and run out on a gel, distinct bands are not easily detectable. Instead, the likely outcome is a broad smear encompassing a large range of fragment sizes. To help resolve distinct bands that can be compared among samples, radioactive DNA probes complementary to the DNA samples are used that hybridize to the digested DNA. In our lab exercise the crime scene DNA and suspect DNA samples originate from plasmids that contain relatively few cut sites. Thus, there is no need to perform the laborious task of Southern blotting. In other words, distinct bands can be easily visualized on a gel following restriction digestion.

Activity:

This will be a **two-week lab exercise** to familiarize you with techniques commonly used in a molecular genetics laboratory. You will be working in **groups of four** during these two weeks. Thus, each lab section should contain four groups of four students. The breakdown of the two weeks is as follows:

Week 8: Background, pipette review, restriction enzyme digestion

Week 9: Gel electrophoresis of DNA digests, analysis, review questions

***Note that nitrile or latex gloves are required. It is up to you to bring these to lab, as your instructor and technicians will not supply them.**

Week 8:

Following the background lecture material, it may be worthwhile to practice your pipetting skills before you begin the lab exercises. The technicians have provided practice cards and plates that you can use to practice as you see fit. **Do not move on to the actual lab exercise until you feel comfortable with pipetting.**

A. Pipetting and Measurements

Types of Micropipettors

Pipettors are made by many different manufacturers and thus all do not look the same. Learning to correctly use one type of pipettor will provide you the knowledge to use others as they share the same method of distributing small volumes. This lab will illustrate the Rainin Pipetman® micropipettors.



The top of the plunger shows the pipettor size for the Pipetman models. Each pipettor has its own volume range and it is **CRITICAL** to use a pipettor only in its proper volume range. The “P-number” represents the maximum volume in μL that the pipettor can measure. Pipettors are more accurate in the *upper* part of their range. For example, 20 μL should be measured with a P20 rather than with a P200. The four pipettor sizes (P10, P20, P200, P1000) used in our lab will measure from 1 μL – 1000 μL as shown below (Figs. 5,6).

Correctly Adjusting the Pipettors

	P10	P20	P200	P1000
Volume Range (μL)	1-10	2-20	20-200 50-200 *	200-1000
	tens 0	tens 1	hundreds 1	thousands 0
	ones 4	ones 7	tens 5	hundreds 8
	tenths 7	tenths 8	ones 4	tens 6
	4.7μl	17.8μl	154μl	860μl

Fig. 5. Gilson Pipetman pipetting ranges chart. Note that the P200 officially has a range from 50-200 μl . Refer to PowerPoint slides for color version.

Tutorial on Proper Usage

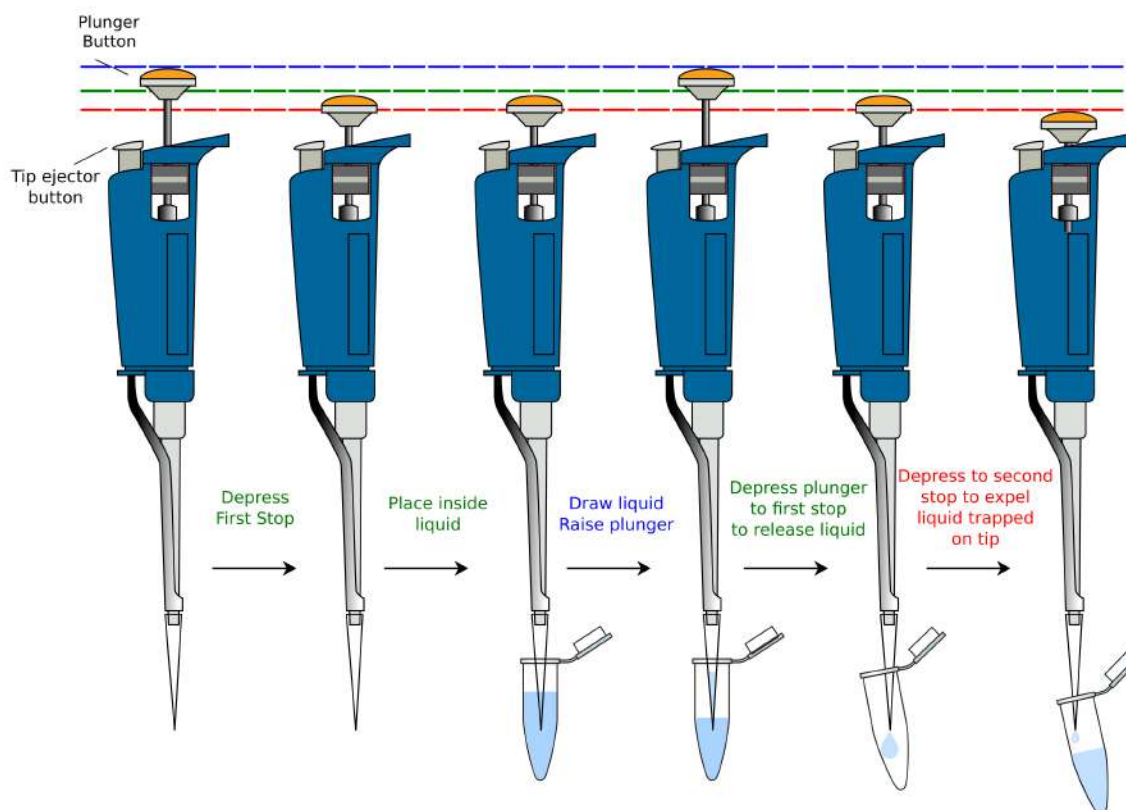


Fig. 6. Pipetting sequence. Ensure the plunger is depressed outside of the liquid to displace air and avoid blowing bubbles into solution. Carefully draw the plunger up slowly and follow the liquid to avoid drawing air. Depress the plunger in destination tube to the first stop. Depress to the second stop if fluid remains in the tip to expel full volume. Refer to PowerPoint slides for color version.

https://www.youtube.com/watch?v=uEy_NGDfo_8

Rules for use of the micropipettors:

These are precision instruments which can easily be damaged. Treat them with respect and care. They are essential for your success in this course and shared amongst numerous students.

1. Never measure higher or lower than the range of the pipettor allows.
2. Never turn the volume adjuster above or below the range indicated on the pipettor or you risk breaking the instrument.
3. Never allow liquid to get into the micropipettor.
 - this causes contamination
 - this weakens the seal on the o-rings and can damage them
4. Never use the micropipettor without an appropriate tip. Never force a tip that doesn't fit.
5. Never invert or lay down the micropipettor with liquid in the tip.
 - liquids will roll into the piston this way
6. Never let the plunger snap back when filling or ejecting liquid.



7. Never immerse the barrel in fluid.
 - this causes contamination
8. Never set the micropipettor on the edge of the bench; this may result in the micropipettor falling or being knocked onto the floor.

Exercise: Pipetting Practice

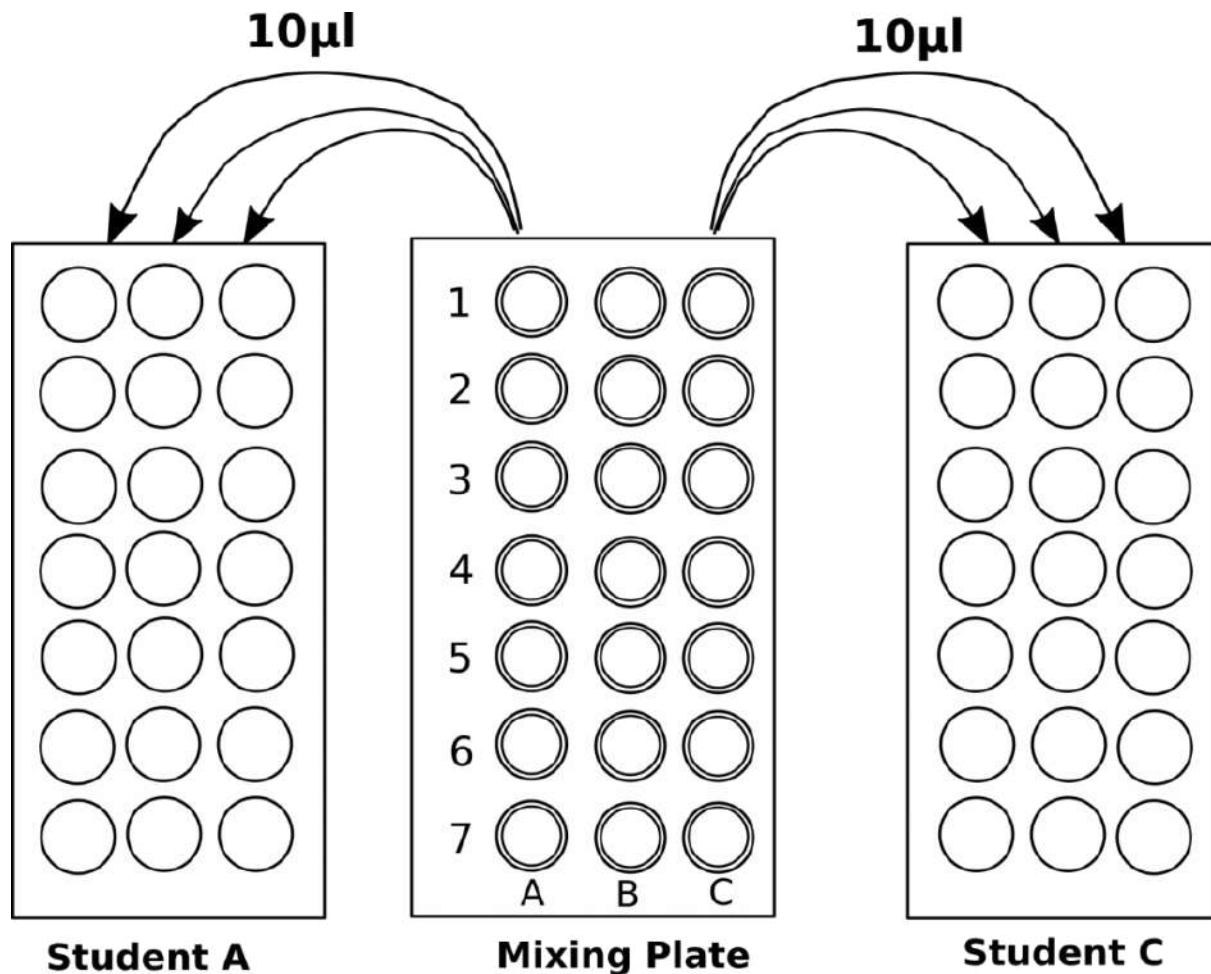
1. Prepare seven dye mixtures as illustrated in the table below.

-One student mixes samples in Column A and a second mixes in Column C

-Column B is left empty and used if one student makes a mistake

2. Each dye mixture prepared in one well to reach a total volume of 45 μl .
3. Pipet 10 μl in triplicate from each well of the mixing plate into the center of the appropriate circles on the target card

Well #	Buffer (μl)	Red (μl)	Blue (μl)	Yellow (μl)	Glycerol (μl)	Alcohol (μl)	Total (μl)
1	40	5	-	-	-	-	45
2	15	-	-	10	10	10	45
3	20	5	10	10	-	-	45
4	30	-	15	-	-	-	45
5	22	13	-	-	-	10	45
6	25	-	10	-	10	-	45
7	20	6	-	6	13	-	45



B. Setting up restriction enzyme digests

Now we are ready to set up the restriction digests. Each group should have one tube labeled **ENZ** that contains the rehydrated restriction enzymes EcoRI and PstI. It important to **keep the restriction enzymes on ice** during the preparation of restriction digests as enzymatic activity decreases at higher temperatures.

1. Label colored tubes as follows:

Green = CS (crime scene DNA)
 Blue = S1 (Suspect 1)
 Orange = S2 (Suspect 2)
 Violet = S3 (Suspect 3)
 Pink = S4 (Suspect 4)
 Yellow = S5 (Suspect 5)
Clear = negative control

Also include your name or initials, date, and lab section on your tubes.

2. The technicians should have already rehydrated both the enzymes and the DNA samples and have transferred 10 ul of the stock DNA into the corresponding colored tube. For your negative control you will use water.
3. Add 10 ul of the enzyme mix to each tube. **Again, make sure to use a fresh pipette tip between samples to prevent contamination.** Pipette the enzyme, sample mixture up and down a few times to homogenize the sample. Close lid on all caps.
4. If you see bubbles in your tubes, gently flick the bottom or gently tap the tube on your lab bench. Make sure the cap is securely closed to prevent spillage. You may also wish to briefly vortex the sample and/or place it in a mini- or microcentrifuge for a few seconds. **When using a microcentrifuge you must make sure it is balanced or you may damage the rotor.**
5. Place the tubes inside a dry bath and incubate for 45 min at 37°C.
6. When the incubation period is complete, store digests in the refrigerator until the following week.

Week 9:

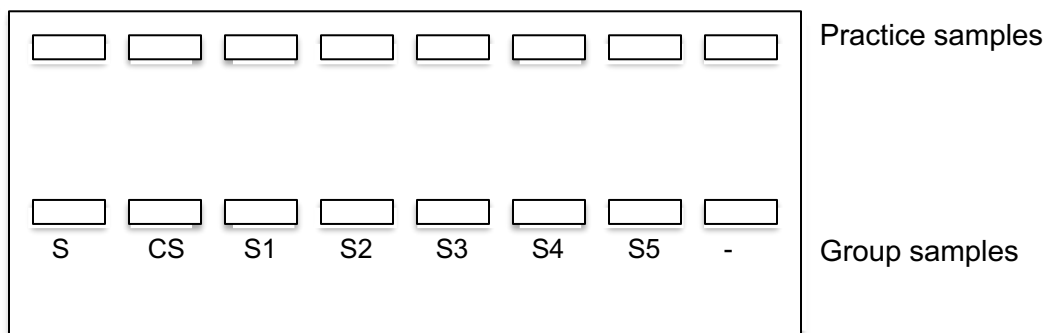
This week we will quantify the results of your restriction digests to determine if any of the suspect DNA samples match the DNA found at the crime scene. Remember that a suspect DNA banding pattern matching the pattern at the crime scene would implicate that individual in the crime.

Each class will prepare four 1% gels in total.

1. For each gel, combine 50 ml of TBE buffer with 0.5 g agarose. Swirl mixture.
2. Microwave mixture for 30 sec – 1 min. Using protective gloves, remove the mixture from the microwave and swirl. **Be careful as the flask will be hot to the touch!** If needed, microwave for another 30 sec – 1 min. The agarose should be completely dissolved in the buffer solution and look clear. Make sure that the solution does not spill out of the flask when boiling.
3. Add 5 ul of SYBR safe solution to each flask and swirl. This is the stain that will help you visualize the DNA in your gel.
4. Gently pour solution into a gel casting tray and place two combs into the gel.
5. Wait for gel to solidify (~20 min).
6. Use one row of wells to practice loading the gel. Try loading 5 ul of dye into each well.
7. Remove your restriction digests from the refrigerator. Make sure there are no bubbles inside the tubes. Gently flick tubes, tap them on the bench, vortex them, or spin them in a microcentrifuge if needed.



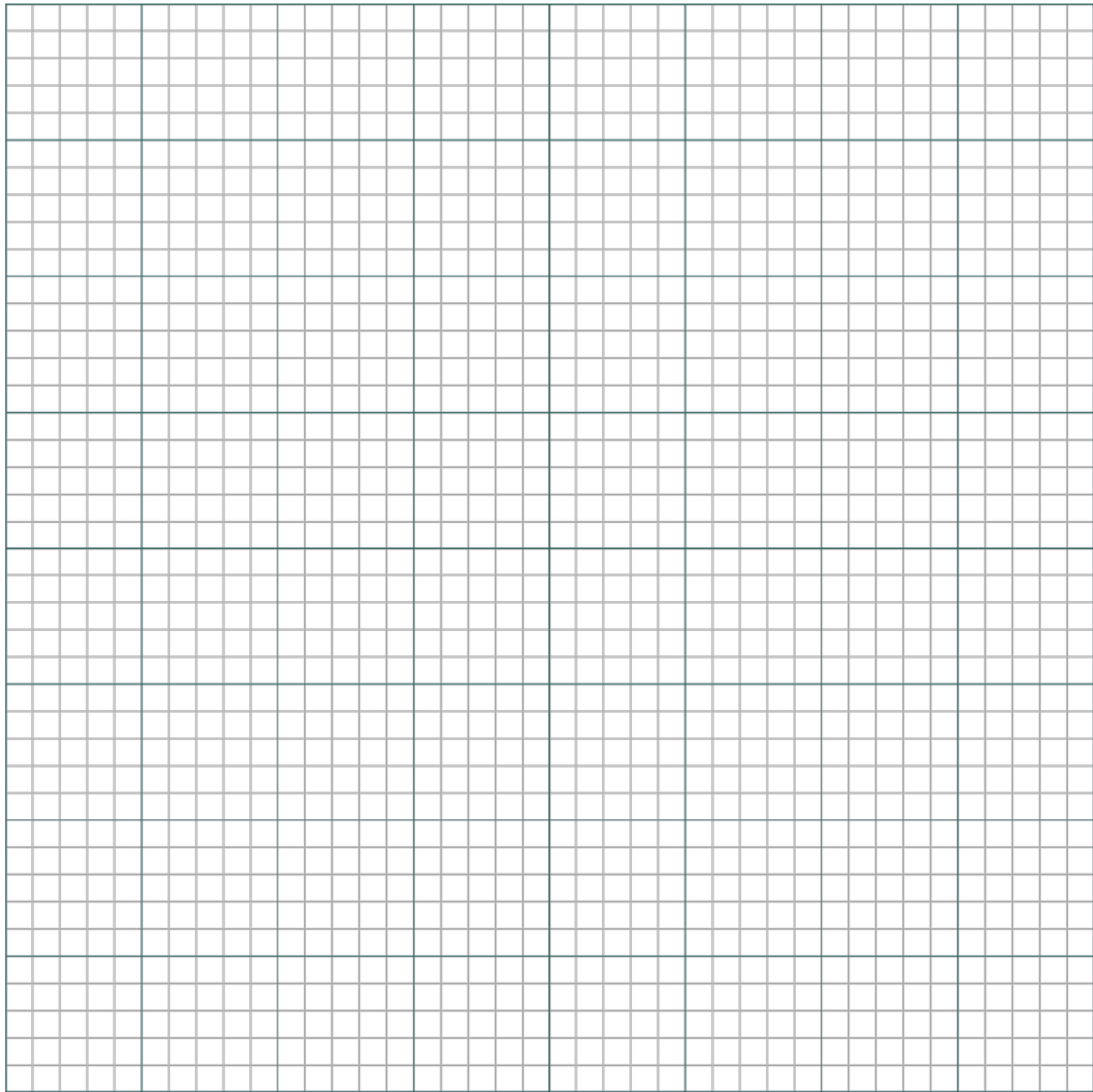
8. Add 5 ul of loading dye to each sample. The loading dye is used to help load your samples into the gel. **Make sure you use a different pipette tip between samples!** Once the loading dye is added, each tube should contain a total of 25 ul (10 ul DNA + 10 ul restriction enzyme + 5 ul loading dye).
9. When solidified, carefully transfer the gel to an electrophoresis chamber. Make sure to fill the chamber with enough buffer to cover your gel. Also, make sure that the wells of the gel are near the black electrode (not the red).
10. Load the gels following the diagram below. **Make sure to use a different pipette tip for each sample when loading.** Use **10 ul of the size standard (S)** and **20 ul of each sample.** Place your negative control in last well.



11. Place the lid on the electrophoresis chamber and plug the correct electrodes into the power supply and the gel rig (black to black, red to red).
12. Turn on power supply and set to 100 V. Electrophorese your samples for ~30-45 min.
13. While waiting, work on the review questions.
14. Turn off the power supply after the loading dye is approximately halfway through the gel. With gloves, carefully remove gel from chamber and place on transilluminator to visualize the results. **Do any of the samples match? Is it possible that any of the suspects committed the crime?**
15. **Use a ruler to measure the distance (mm) of each band from the well. Determine the approximate size of the band (in bp) by comparing it with the Invitrogen 1kb ladder. Fill-in the table on the next page. Use the table to plot the data in the following plot. Your x-axis should be distance (mm) and the y-axis should be fragment size (bp). Be sure to label the axes! What is the general shape of the curve? How can the figure be used to estimate the size of an unknown fragment?**

	Size Standard		Crime Scene Sample		Suspect 1		Suspect 2		Suspect 3		Suspect 4		Suspect 5	
Band	Distance (mm)	Actual size (bp)	Distance (mm)	Approx size (bp)	Distance (mm)	Approx size (bp)	Distance (mm)	Approx size (bp)	Distance (mm)	Approx size (bp)	Distance (mm)	Approx size (bp)	Distance (mm)	Approx size (bp)
1														
2														
3														
4														
5														
6														





Review Questions

1. If a restriction enzyme digested a linear piece of DNA in two places, how many fragments would result? What if the DNA was circular instead of linear?
2. In gel electrophoresis, why is the DNA loaded near the negative end (cathode) of the chamber?
3. How does this experiment compare and contrast to a typical RFLP DNA fingerprinting assay?
4. Describe the role of restriction enzymes in constructing a recombinant plasmid.
5. Which of the suspect DNA samples contained the largest DNA fragment? What about the smallest?

6. A more modern molecular technique to RFLP fingerprinting is called Amplified Fragment Length Polymorphisms (AFLPs). In AFLP analysis, restriction enzymes are again used to digest genomic DNA into multiple fragments. Next, adapters complementary to restriction site overhangs are ligated to the fragments using an enzyme called DNA ligase. These adapters are complementary to primers used to amplify the fragments using the polymerase chain reaction (PCR). Can you think of any potential benefits of AFLP analysis over RFLP? Explain your reasoning.

7. Using what you now know about restriction enzymes, why do you suppose that wells in the gel contained more than one band?

8. How would the voltage used change the rate of DNA electrophoresis? Can you think of any benefit of running the samples at a slow rate?

9. What is the purpose of including a negative control in your experiment? Would DNA bands in the negative control influence your conclusions? Explain your reasoning.

10. What would your gel look like if the DNA were not fragmented?

Molecular Genetics, PCR, and Genotyping

Background:

For the next three weeks we will be performing two molecular genetics experiments to introduce you to some commonly used laboratory techniques to assay genetic variation at the molecular level. All of these experiments rely on a technique called **Polymerase chain reaction (PCR)**. PCR has revolutionized the study of genetics, helping to provide insight in a wide variety of disciplines ranging from medical genetics to ecology and evolutionary biology. The purpose of PCR is to exponentially amplify a specific gene or fragment of DNA, which is subsequently called an **amplicon**. Figure 1 highlights the generally procedure of PCR. **Note that gloves are required throughout all phases of these experiments.**

The general procedure for PCR is as follows:

Polymerase chain reaction - PCR

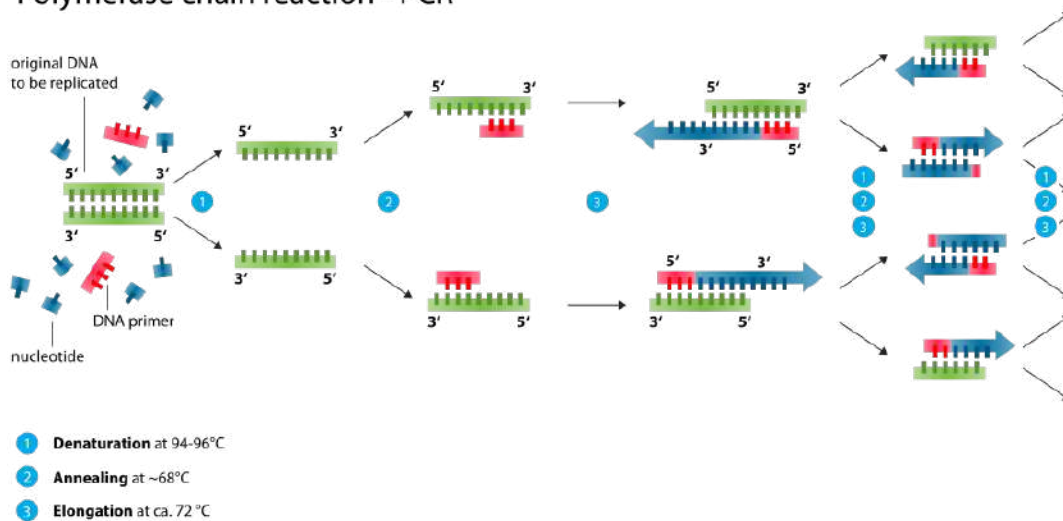


Fig. 1. Graphical depiction of the Polymerase chain reaction (PCR) technique. Green fragments represent the template DNA to be amplified. Primers are shown in red, and amplicons are shown in blue. *Image credit: CC BY-SA 3.0 Enzoklop (author).*

1. Determine the fragment of DNA to be amplified. This is usually between 500-1000 bp.
2. Design or obtain primers flanking the region of interest that are complementary to the template DNA. Primers are needed to 'prime' the reaction as DNA polymerase can only extend DNA in a 5' – 3' direction.

3. Create a **master mix** containing all the needed reagents. Master mixes usually contain **water, PCR buffer, dNTPs (nucleotides), both primers, taq DNA polymerase**, and the **template DNA**.
4. Reactions are usually performed in a machine called a **thermal cycler**. Thermal cyclers automate the cycling of different temperatures that are needed to amplify the DNA. Multiple cycles are generally performed (e.g. 35), with each cycle consisting of three steps: denaturation (95 °C), primer annealing (varies, generally between 50 °C – 60 °C), and extension (72 °C). In addition, there is usually a final extension phase and an initial denaturation phase.
5. When the PCR is complete, we can use gel electrophoresis to determine if the reaction worked. In many cases, we expect to see a single band in each well, particularly if we amplify a haploid locus (e.g. mtDNA, cpDNA). Conversely, in some cases individual alleles may be seen in a diploid marker, but this requires that the alleles are quite different in size and/or a polyacrylamide gel is used instead of agarose.
6. The next step is to purify the PCR product to remove primers and unincorporated nucleotides. This can be done via spin column techniques or with enzymatic treatment.
7. The purified PCR products can then be used for downstream applications such as direct DNA sequencing.

Lab overview and objectives:

Over the course of three weeks we will be performing two sets of experiments to introduce students to fundamental concepts in molecular genetics (Fig. 2). Upon completion of these labs students should be able to explain how restriction enzymes work, discuss what PCR is and how it is commonly used in molecular biology and genetics, understand why gel electrophoresis is performed, and demonstrate basic knowledge of Sanger sequencing and scoring of genetic polymorphisms.



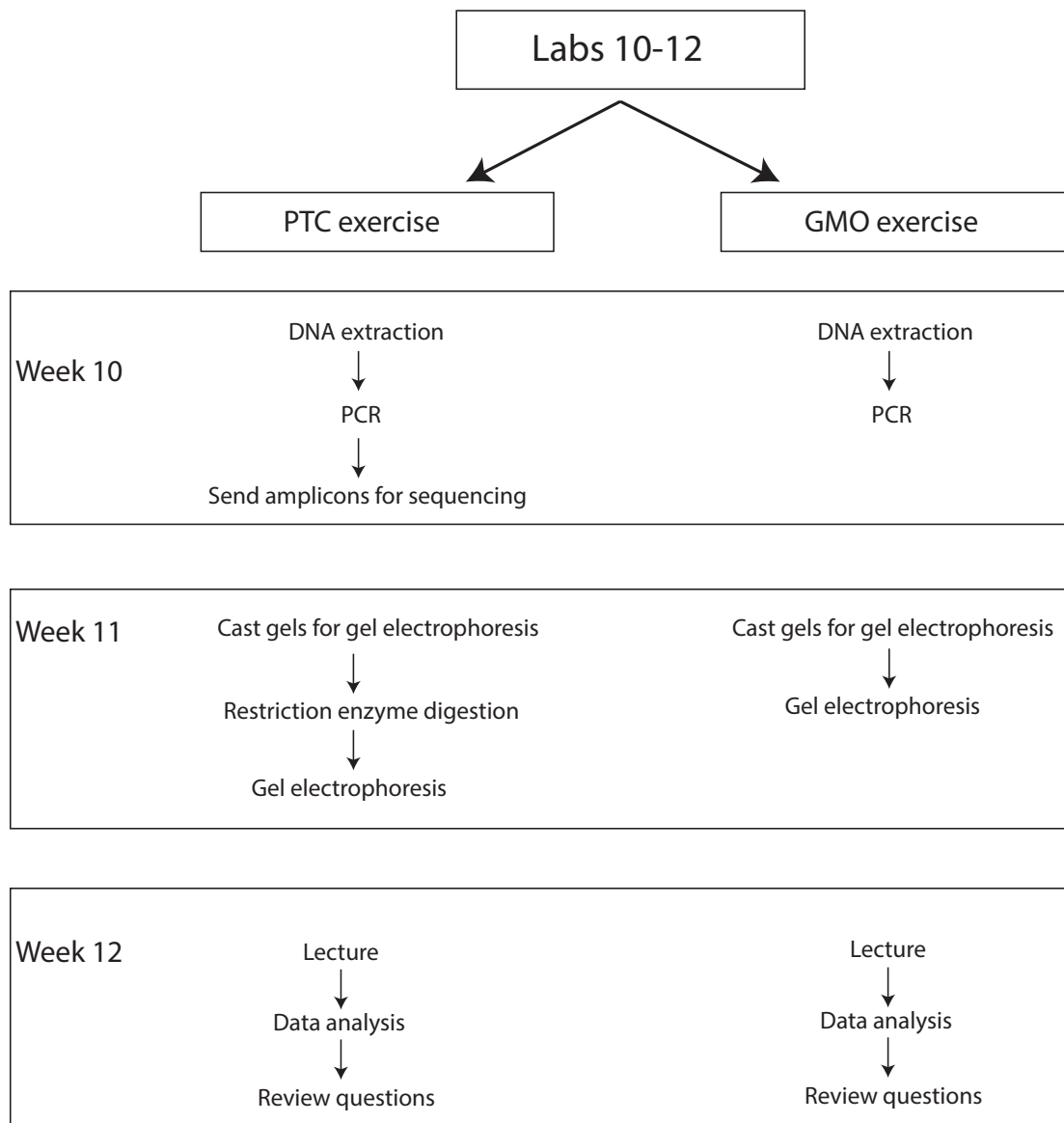


Fig. 2. General overview of lab activities for Weeks 10-12. Each week students will simultaneously work on both the PTC and GMO activities.

Experiment A: PTC Genetics

As stated above, we will be performing two different experiments over the next three weeks. The first experiment focuses on the molecular genetics of a taste receptor found on the tongues of vertebrates. Many animals have specific taste receptors that help determine if a substance is bitter. This is often useful, as bitter substances are generally correlated with some level of toxicity. For this lab, we will be working with a **synthetic, non-toxic chemical called phenylthiocarbamide (PTC)** to mimic naturally occurring bitter compounds.

Humans can be classified along a spectrum of tasting ability for PTC. Some people detect a strong bitterness to the substance, whereas others taste nothing. The receptor protein is encoded by an autosomal gene called *TAS2R38*, which contains two common alleles and several uncommon variants. Differences in the two common alleles are due to a **single nucleotide polymorphism (SNP)** that changes one of the amino acids. The tasting allele, *T*, is thought as dominant to the non-tasting allele, *t*. Homozygous dominant (*TT*) individuals generally describe the taste as very bitter, heterozygotes (*Tt*) may detect mild bitterness, and homozygous recessive individuals (*tt*) will not detect any bitterness. Although many researchers treat the genetics of PTC tasting as a simple Mendelian trait (one gene, two alleles, complete dominance), the actual genetics is probably more complex involving many genes and also including an environmental component. Thus, the ability to taste PTC is likely a **polygenic trait**. Even at the single gene level (*TAS2R38*), there is no consensus as to whether the trait exhibits complete versus incomplete dominance.

Since having the tasting allele, *T*, would seem like a benefit due to an organism's ability to detect potentially toxic compounds (and avoid or adapt to them), why has natural selection not removed the non-tasting allele, *t*, from populations? One reason might be **balancing selection** due to **overdominance** (heterozygote advantage). Balancing selection tends to maintain multiple alleles in a population. In the case of PTC, there may be some selective advantage of possessing the non-tasting allele in the heterozygote form (similar to the case of sickle-cell anemia). Clearly, more research needs to be conducted to determine the potential benefit of the non-tasting allele.

You will use PCR to amplify a **221 bp fragment of the PTC gene** to determine the genotypes corresponding with the different phenotypes (tasting ability). Before beginning the molecular component of this lab we will quantify the phenotypes of the class in the table below. Each person should taste the control paper first to make sure no bitterness is detected.

Your PTC phenotype: _____

Class Data

Phenotype	Number	Frequency
Strong PTC Taster		
Weak PTC Taster		
Non-Tasters		
Total		1.0

Based on the results in the table above, does it appear that PTC tasting ability follows simple Mendelian inheritance patterns? Explain your reasoning.



Procedure:

A general overview of the PTC laboratory procedure is depicted in Fig. 3. Note that allelic differences are due to a G-C polymorphism at base 145 that changes the amino acid from proline (taster) to alanine (non-taster). **Use the gel diagram below to draw predicted banding patterns for each of the three digested genotypes and an undigested amplicon. The ladder well can be used to indicate bands of known size.**

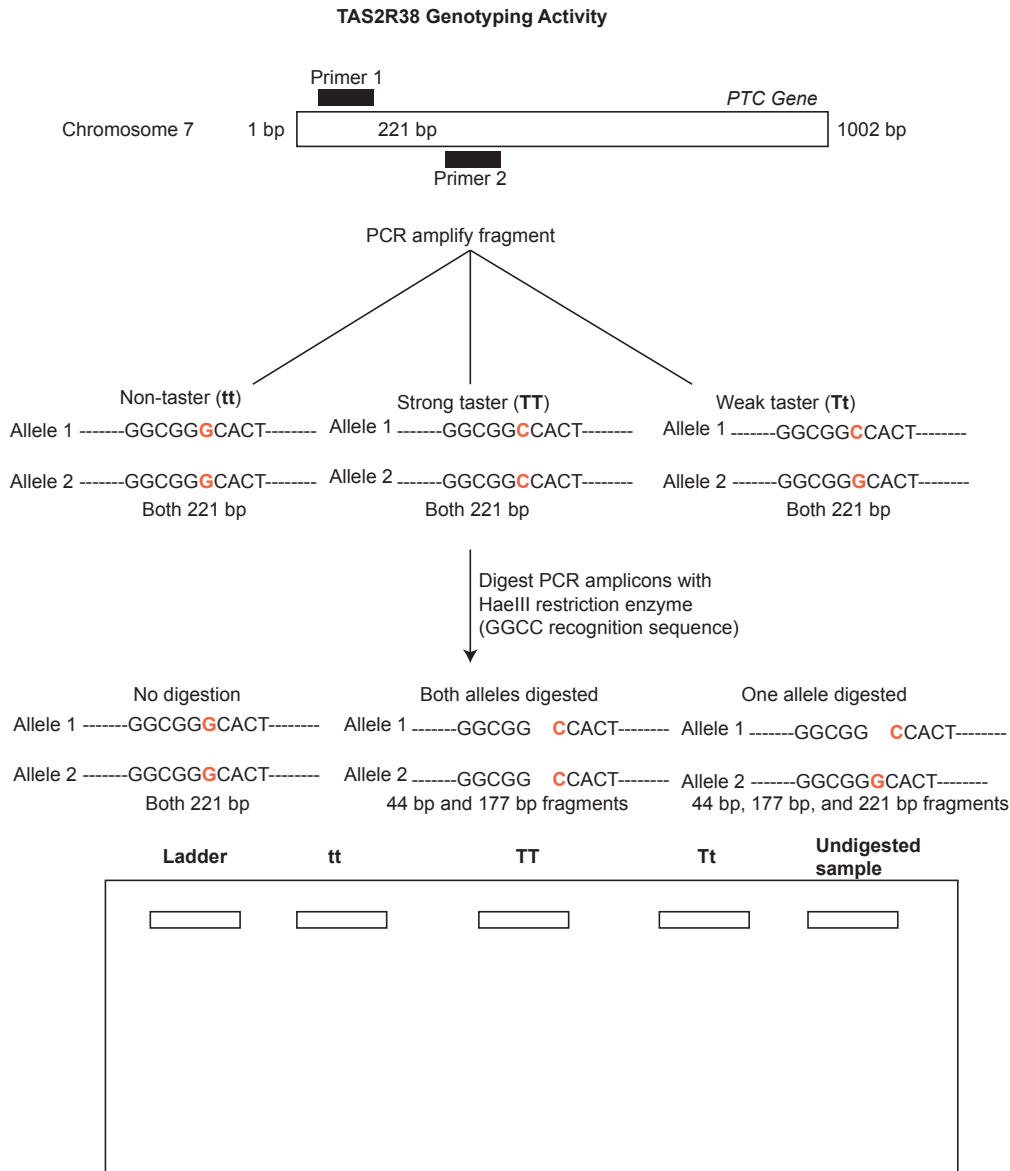


Fig. 3. Simplified workflow for quantifying allelic polymorphism at the TAS2R38 gene.

PTC Genetics Lab Schedule

Week 10 – DNA extraction from cheek cells, set-up PCRs and send amplicons for sequencing

Week 11 – Restriction enzyme digestion, gel electrophoresis

Week 12 – Lecture, data analysis, and review

Day One (Week 10) - Using Cytobrush to extract DNA from cheek cells

1. Use sterile cytobrush and insert into mouth.
2. Brush cytobrush on inside of cheek 25 times.
3. Swirl cytobrush in 100 μ l of Chelex suspension (10% w/v).
4. Place centrifuge tube with Chelex and cell suspension on 100 °C heat block for 10 minutes.
5. Centrifuge tubes at maximum speed for 5 minutes.
6. DNA is in the supernatant (avoid beads at bottom).
7. Store DNA in -20 °C.

Day One (Week 10) - PCR with PCR beads and DNA sequencing

1. Add 22 μ l of primer mix (forward, reverse and loading dye) to PCR tubes containing PCR beads.

Primers:

5'-CCTTCGTTTTCTTGGTGAATTTTTGGGATGTAGTGAAGAGGCCGG-3' (Forward Primer)

5'-AGGTTGGCTTGGTTTGCAATCATC-3' (Reverse Primer)

2. Ensure that the bead is dissolved.
3. Add 3 μ l of DNA.
4. Load samples on the thermal cyclers and run the “**PTC**” program.
5. Store amplified product at -20 °C until you are ready for the next stage of the experiment.

When the PCRs are complete, your instructor will send your amplicons for DNA sequencing using the **Sanger sequencing** technique. An overview of the Sanger method is provided in Fig. 4. Briefly, the PCR amplicon is combined with one of the primer sequences (forward or reverse), DNA polymerase, dNTPs and ddNTPs and cycled in a thermal cycler at certain conditions. During primer elongation, a ddNTP (instead of a dNTP) can become added to the growing sequence. The addition of a ddNTP causes termination of elongation, which is why Sanger sequencing is also called chain termination. Sequencing fragments are then separated by gel or capillary electrophoresis and a laser is able to differentiate between

individual bases to produce an electropherogram. The reverse primer will be used for sequencing in this experiment.

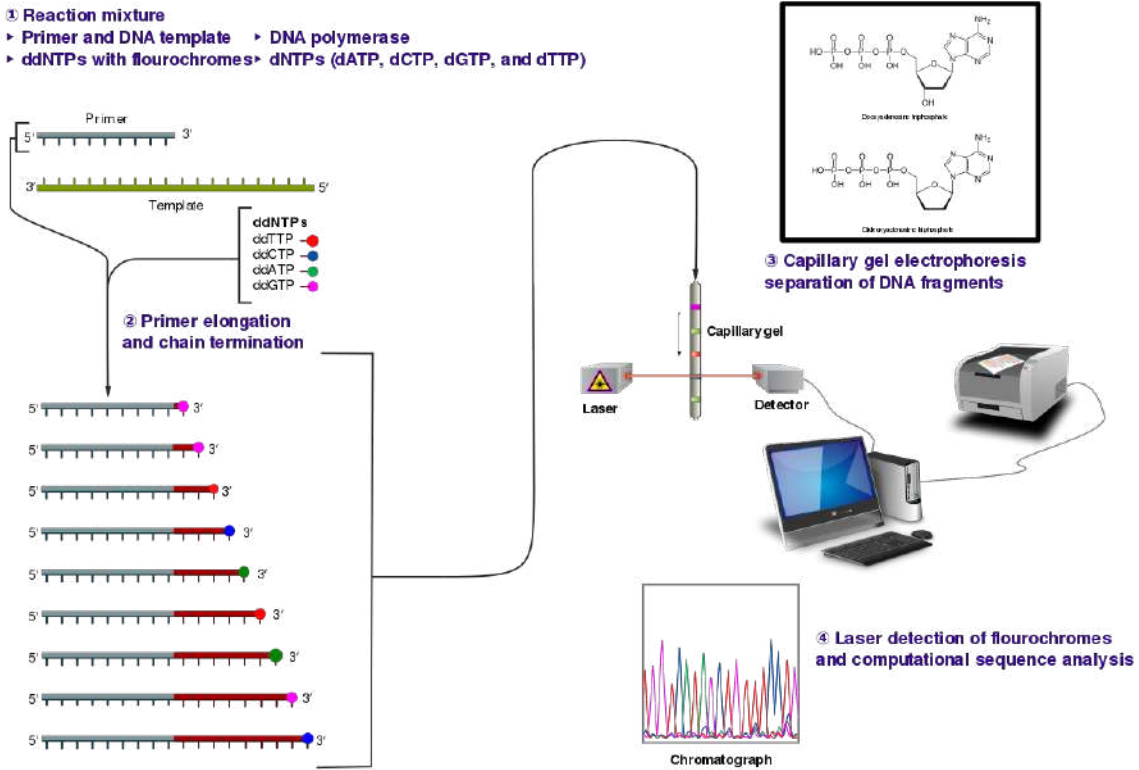


Fig. 4. Overview of the Sanger method of DNA sequencing. Note that the addition of a ddNTP (instead of a dNTP) causes the termination of strand synthesis. Fragments are then sorted by size and a laser is used to determine the sequence of bases (nucleotides). *Image credit: CC BY-SA 3.0 Estevezj (author).*

Day Two (Week 11) - Restriction enzyme digestion and gel electrophoresis

Our next goal is to setup restriction digests on the PCR products to quantify allelic polymorphisms. Recall that the recognition site for the *Hae*III enzyme is GG|CC. In other words, the enzyme will only cut the DNA at these sequences.

Gel Casting (Gels will be used for both laboratory experiments)

1. Each group of four (4) will prepare a 2% agarose gel.
2. In a flask, combine 50 ml TBE buffer with 1 g agarose. Gently mix the solution.
3. Microwave the buffer-agarose mixture at 30 sec intervals, mixing the solution between heating. **Make sure you wear protective gloves as the flask will be hot!** The agarose should be completely dissolved in the buffer and be clear.
4. Add 10 ul SYBR Safe DNA stain to the hot flask and gently mix.
5. Place two combs into each gel apparatus. **Make sure the comb with 10 wells is placed facing down.** Let the hot gel mixture cool a bit before pouring (~3 min). Carefully pour the gel into the gel tray, making sure that no liquid is spilled.
6. Gels may be placed in the refrigerator to expedite solidification. This should take no more than 20 min.

DNA digestion

1. Each student should obtain two new microcentrifuge tubes. Label one tube with your initials and the letter 'U' for undigested and one tube with the letter 'D' for digested.
2. Transfer 5 ul of PCR product into tube labeled 'U' and 5 ul into tube labeled 'D'.
3. Add 1 ul of *Hae*III enzyme to **tube 'D'**.
4. Briefly vortex or pipette-mix the mixture to homogenate.
5. Incubate for 10 min at 37 °C.
6. **Save the remaining 15 ul of PCR product for DNA sequencing.**

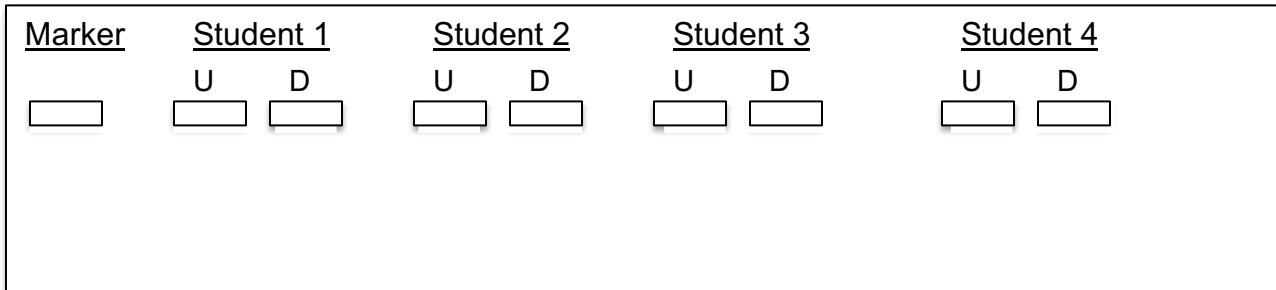
Gel electrophoresis

We will use gel electrophoresis to separate and visualize the DNA fragments resulting from digestion with the *Hae*III restriction enzyme. Your goal is to compare the observed genotypes among individuals.

1. When adequately solidified, place gel in gel apparatus. Make sure that the wells of the gel are near the negative (black) electrode.



- Each student should load two samples into the gel—one undigested sample and one digested with the restriction enzyme. Each gel should also include a DNA ladder to help size the fragments (load **5 ul** of the DNA ladder). Below is a representative diagram of how the gels should be loaded for one group. U = undigested; D = digested.



- Run gel for 20 min at 120 V.
- Visualize results using transilluminator and obtain genotypes. Pay particular attention to your own banding pattern.
- Use your smartphone to take a picture of your gel.

Day Three (Week 12) – Lecture, data analysis, and review questions

- Obtain your DNA sequence from your instructor and analyze it using the software **FinchTV**. Specifically, import the .ab1 files into the program. Use the chromatograms to edit the sequences (e.g. trim low quality bases from the beginning and end of reads). Low quality bases in the middle of sequences can be replaced by an 'N'. Once the sequence has been edited, export it as a new .seq file. Take some time editing sequences from other students as well. What nucleotide is present at ~position 146? What does this indicate? Note that a heterozygous site is indicated by two peaks with relatively even intensity. Use Fig. 3 to help determine your genotype.
- Compare and contrast your DNA sequence to your banding pattern on your gel. Do both methods yield the same results? Can you unambiguously determine if you are homozygous or heterozygous?

Review questions:

1. Draw your banding patterns in the gel image below.

<u>Marker</u>	<u>Student</u>	
	U	D
<input type="text"/>	<input type="text"/>	<input type="text"/>

2. What is your genotype? How does your genotype correspond to your phenotype?
3. Are you homozygous or heterozygous? How do you know?
4. Fill-in the table below regarding observed genotypes for the entire class. How do the genotype frequencies compare with the phenotype frequencies documented in the previous table?

Class Data

Genotype	Number	Frequency
<i>TT</i>		
<i>Tt</i>		
<i>tt</i>		
Total		1.0

5. Use the information in the table to calculate the frequency of the two alleles.

6. Use the Hardy-Weinberg (HW) equation to determine if the class data deviate significantly from HW expectations. Remember that significance is often assessed by using the **chi square test with 1 df** (when using two alleles). The critical value according to the table is 3.841 using an alpha of 0.05. The HW formula is as follows:

$$p^2 + 2pq + q^2 = 1$$
$$p + q = 1$$

7. If diploid species possess two alleles per locus, how do you explain the banding patterns seen in the samples?
8. In your own words, discuss the benefits of possessing the tasting allele. After all, phenylthiocarbamide is not a naturally occurring compound. Why might the non-tasting allele be maintained in natural populations?

9. How do your genotyping results correspond to your DNA sequence? Remember that in this experiment the two alleles are distinguished by a **G-C substitution** at base ~146. What would you expect the sequence of a heterozygote to look like? Explain your reasoning.
10. Explain how restriction enzymes work and how *HaeIII* was able to discriminate between the tasting and non-tasting alleles.
11. Discuss how PCR works and how it can be used to clonally amplify specific fragments of DNA.

Experiment B: Genetically Modified Organisms (GMOs)

There is continual interest in the potential uses of genetically modified organisms (GMOs) to benefit both humans and natural ecosystems. This is particularly true with GMO crop species. By inserting foreign genetic material into species of agricultural importance, scientists can create strains that are pesticide- and herbicide-resistant, produce larger fruits, and exhibit particular nutrient composition in their species of interest. Unfortunately, creating GMO crops is an expensive and laborious process. Furthermore, there is considerable debate as to the pros and cons of creating GMOs. For example, opponents argue that GMOs will lead to the evolution of herbicide-resistant weeds and that there may be negative health outcomes from consuming GMO species.

The first step to create a GMO is to identify a gene and trait of interest. As an example, genes from the bacterium *Bacillus thuringiensis* are commonly inserted into crop species because many of these genes encode proteins that serve as a pesticide to certain insect species. Thus, farmers will not have to spray harmful (and carcinogenic) synthetic pesticides on crops, which may have a substantial influence on human health. The next step is to isolate, or clone, the gene of interest into a DNA vector. A DNA vector is a circular piece of DNA (usually a bacteria plasmid) that can be genetically engineered based on its proposed usage. In addition to cloning the gene of interest, other fragments of foreign DNA are also inserted into a vector. These include regulator sequences such as a 35S promoter region from the cauliflower mosaic virus (CaMV 35S) and a terminator region from the nopaline synthase (NOS) gene from the bacterium *Agrobacterium tumefaciens*. These regulatory elements are commonly used in GMOs to maintain high levels of gene expression in nearly all the cells of the plant. Next, researchers must find a way to introduce the foreign DNA into plant cells. It is generally not possible to transform all the cells of a grown plant. Instead, scientists isolate a few plant cells that will incorporate the exogenous DNA. Multiple techniques can be used for transformation including infection with bacterial species, using a 'gene gun' to bombard cells with the DNA, or using a technique called electroporation to create pores in plant cell membranes. Regardless of the technique, in order to be effective the foreign DNA must pass through the cell wall, plasma membrane, and nuclear membrane to be successfully incorporated inside the plant cell genome. The next step is to determine which plant cells have taken up the foreign DNA into their genome. To aid in this endeavor, additional selectable markers are often used. Finally, researchers will add hormones to the plant cells to stimulate cell division, which will eventually yield an entire plant composed of genetically modified cells.

For this lab exercise, you will test common grocery items to determine if they are genetically modified. Approximately 85% of GMOs contain either the CaMV 35S promoter or the NOS terminator sequence, both of which are ~200 bp long. **Common GMOs in the United States include corn and soy-based products.** In addition, many strains of papaya often contain GMO sequences that can be determined using molecular genetic techniques. Wheat and rice products are generally lacking genetic modification. Throughout this exercise, multiple controls will be used to make sure that the experiments runs adequately. This is explained in greater detail below.



As the molecular techniques in this exercise are similar to those used in the PTC genotyping lab, we will work on both exercises simultaneously over the course of **three weeks (Weeks 10-12)**. The GMO lab will be divided as follows:

GMO Lab Schedule

Week 10 – DNA extraction from food samples and PCR

Week 11 – Gel casting and electrophoresis

Week 12 – Lecture, data analysis, and review

Day One (Week 10) – DNA extraction from food samples

You were asked to bring in food items from the grocery store to test for evidence of a GMO. A class of 16 should be divided into four groups of four. There should be enough supplies for each group to test a different food item.

1. Obtain two 1.5ml microcentrifuge tubes. Label one as “non-GMO” and the second as your test food.
2. Place a crumb of “non-GMO” (that comes with the lab) or your test food into the corresponding tube.
3. Add 50 ul dH₂O to each tube.
4. Twist a clean plastic pestle against the inner surface of the tube for 1 min to forcefully grind the crumb. Be sure to use a different pestle for each sample.
5. Add 300 ul of InstaGene matrix to each tube, pipette up and down a few times to homogenize the ground slurry with the InstaGene matrix.
6. Place each tube (non-GMO control, test sample) in a dry bath and incubate for 5 min at 95 °C.
7. Place tubes in a microcentrifuge and spin for 5 min at maximum speed to separate the DNA, which will be in the supernatant following centrifugation.

Day One (Week 10) – Set-up PCRs

Following DNA extraction we will set-up our PCRs. Remember that we will be using PCR to determine if the CaMV 35S sequence is present that would indicate that the item has been genetically modified. **Each of the four groups will set-up six (6) reactions.** Use the following number scheme to label your tubes. Also be sure to include your initials, the date, and the lab section on the tubes. **Make sure you use different pipette tips between samples!** Also make sure that you do not disturb the InstaGene beads when transferring your DNA. When instructed, place your samples in the thermal cycler and execute the program “GMO.”



Tube	Primer Mix (forward, reverse, loading dye)	DNA
1	22 ul Plant Primer Mix	3 ul non-GMO control DNA
2	22 ul GMO Primer Mix	3 ul non-GMO control DNA
3	22 ul Plant Primer Mix	3 ul test food DNA
4	22 ul GMO Primer Mix	3 ul test food DNA
5	22 ul Plant Primer Mix	3 ul GMO positive control DNA
6	22 ul GMO Primer Mix	3 ul GMO positive control DNA

You will notice that each DNA sample is being amplified twice using two different master mixes, one plant and one GMO. The plant master mix contains primers that will amplify ~200 bp of a gene within the chloroplast DNA (cpDNA). Why do we include primers for this gene? This helps us to guard against **false negative** results. For example, we would be fairly confident that a food item did not contain GMO sequences if no bands were detected using the GMO primers, but the cpDNA band was detected. Conversely, if no bands were detected with either master mix we cannot be entirely confident in our results that the item is non-GMO. The lack of bands with both master mixes might be due to a poor DNA extraction.

Looking at the table, can you provide a reason why we are including non-GMO control DNA in the experiment? How does this make the experiment more reliable?

You will also notice that our DNA samples contain a GMO positive control. Discuss why this is useful to our experiment. If the GMO positive control yielded no bands with the GMO master mix, what might you conclude?

In general, why is it useful to include multiple controls when performing molecular genetic experiments? How might this influence the diagnosis of various human diseases?

Day Two (Week 11) – Gel casting and electrophoresis

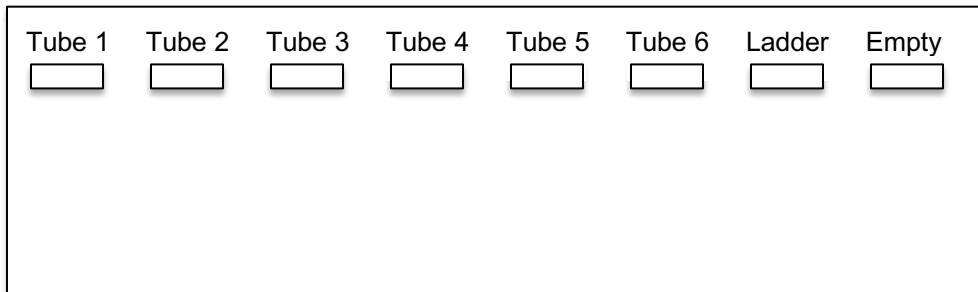
This week we will be performing gel electrophoresis to determine the banding patterns of your food samples to determine if any are genetically modified. Remember, about 85% of GMOs will possess the CaMV 35S sequence, which is ~200 bp in length. The cpDNA gene is also ~200 bp.

Gel Casting

1. **See directions above for PTC exercise. Each group's gel will contain samples from the PTC exercise AND the GMO exercise!**

Gel electrophoresis

1. Place gel into the electrophoresis chamber. Make sure there is enough 1X TBE buffer to cover to gel.
2. Load **20 μ l** of each PCR sample (and **5 μ l** of the DNA ladder) into the gels following the diagram below. Refer to the table above for tube-sample correspondence.



3. Run gel for 20 min at 120 V.
4. Visualize results using transilluminator and obtain genotypes.
5. Use your smartphone to take a picture of your gel.

Day Three (Week 12) – Lecture, data analysis, and review questions

1. Draw your group's banding patterns obtained following gel electrophoresis in the image below.

Tube 1	Tube 2	Tube 3	Tube 4	Tube 5	Tube 6	Ladder	Empty
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Review questions:

1. Why did we include primers to amplify a cpDNA gene in this experiment?
2. Which foods tested showed signs of GMOs?
3. Why were 2% gels used in the GMO experiment?

4. How were you able to determine that a food sample contained GMOs?

5. Suppose you tested a food item that you knew was genetically modified, but the results of this experiment were negative. How might you explain the results?

6. Plant cells have how many genomes? What are they? Which genomes were examined in this study?

7. Why did your instructor recommend that the non-GMO food control sample be prepared using the mortar and pestle prior to the test food sample?

8. What reagents are used in a typical PCR? What is the function of each component?



9. Explain why DNA ladders are usually included during gel electrophoresis.

10. One aspect of PCR that can be modified is the annealing temperature. In general, higher annealing temperatures show more specificity towards a single template, whereas lower annealing temperatures show less specificity and may bind to multiple regions throughout the genome. Discuss how using an annealing temperature that is too high or too low might influence the results of a PCR assay (and gel electrophoresis results) such as the one used in this study.



Introduction to DNA Barcoding and Evolutionary Genetics

Background:

For the next two weeks we will be performing a DNA barcoding exercise followed by a brief introduction to bioinformatics and evolutionary genetics. So what is DNA barcoding exactly? In short, DNA barcoding uses molecular genetic tools to aid in species identification and delimitation. It can also be used to help determine the presence of 'cryptic' species unknown to science. Many times in nature members from two species may look similar morphologically, but exhibit significant differences in molecular sequences. Thus, we may say that one of the two species is cryptic, meaning that we cannot observe differences based on visual examination. The process of DNA barcoding is analogous to a barcode you might find on a grocery store item. When the food barcode is scanned at the register the machine tells you what it is. Similarly, when a DNA sequence is obtained from an unknown specimen it can be compared to a database to aid in identification.

DNA barcoding has received tremendous interest over the years as threats to global biodiversity continue to increase. Several websites are devoted to collecting barcode sequences from all species on Earth (e.g. <http://ibol.org/>, <http://www.barcodeoflife.org/>). For all organisms, partial sequences of a single gene are used as a barcode. For most animals the gene commonly used is ~650 bp of the mitochondrial gene cytochrome c oxidase subunit I (COI or *cox1*). Unfortunately, this gene is not a useful barcode for other eukaryotes such as plants and fungi. For these groups other genes must be used (nuclear ITS for fungi, chloroplast *rbcL* and *matK* for plants). Below is the general procedure used for DNA barcoding.

1. Obtain fresh tissue samples from specimens of unknown identity.
2. Extract genomic DNA from tissue samples and use PCR to amplify a barcode region (e.g. COI).
3. Sequence purified PCR amplicons.
4. Edit raw sequences, and compare cleaned sequences to database to obtain taxonomic identity. A commonly used database is GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).

For this lab we will be performing DNA barcoding on fish samples from your local grocery store, fish market, or sushi restaurant. Many times, retailers will mislabel fish to try to market their product as a more exotic/expensive species. As many fish taste similar, the majority of people are unable to discern any difference. It is presently unknown at which stage the fraudulent activities actually occur, although the commonness of the phenomenon is likely having strong negative ecological and economic impacts. DNA barcoding can be used to help determine the species identity of the fish being consumed. Commonly mislabeled fish include **Alaskan/Pacific cod, Alaskan/Pacific halibut, salmon, sea bass, red snapper, lemon sole, white tuna, grouper, and walleye** (Warner et al. 2013; Arnold et al. 2017).

Lab breakdown:

Week 13: Collect fish sample from local vendor, extract DNA, setup PCRs.

Week 14: Gel electrophoresis, bioinformatics, and phylogenetic analysis

Week 13:

Tissue collection

Each student is expected to bring in a small sample of fish tissue for the lab. A small square sample about the size of a sugar cube is sufficient to obtain high quality DNA. The sample should remain frozen until it is brought to class. Students can wrap the tissue in aluminum foil to bring to class. See above for recommendations regarding species to target. **Remember which species the fish was sold as!**

DNA extraction

1. Place sample in a clear 1.5 mL microcentrifuge tube.
2. Add 100 μ l of nuclear lysis solution to tube. Twist a clean plastic pestle against the inner surface.
3. Add 500 μ l additional nuclear lysis solution to tube.
4. Incubate the tube in a water bath or heat block at 65°C for 15 min.
5. Centrifuge for 4 min at maximum speed to pellet proteins and cellular debris.
6. Transfer 600 μ l of supernatant to a clean labeled tube.
7. Add 600 μ l isopropanol.
8. Centrifuge for 2 min at maximum speed to pellet DNA.
9. Pour off supernatant and add 600 μ l of 70% ethanol to wash the pellet.
10. Centrifuge tube for 2 min at maximum speed and carefully remove the solution.
11. Air dry the pellet for 10 min and add 100 μ l DNA rehydration solution (TE).
12. Incubate DNA at 65°C for 5-10 min to dissolve.

PCR amplification of ~700 bp of COI

1. Add 22 μ l of primer mix (forward, reverse and loading dye) to PCR tubes containing PCR beads.

Universal fish primers:

5'-TCA ACC AAC CAC AAA GAC ATT GGC AC -3' (FishF1)

5'-TAG ACT TCT GGG TGG CCA AAG AAT CA -3' (FishR1)

2. Ensure that the bead is dissolved.
3. Add 3 μ l of DNA.



4. Load samples on the thermal cyclers and run the “**FISH**” program. This program uses the following cycling conditions: initial denaturation at 95°C for 15 min, 35 cycles of denaturation at 95°C for 30 sec, annealing at 54°C for 30 sec, extension at 72°C for 1 min, followed by a final extension step at 72°C for 10 min with samples held indefinitely at 4°C. ****Make sure to label your PCR tubes clearly and properly!**
5. Store amplified product at -20 °C until you are ready for the next stage of the experiment. Your lab technician will remove the samples from the thermal cycler when finished. He/she will also take an aliquot to ship for sequencing so the results will be ready during the next lab.

Week 14:

Gel casting and electrophoresis

Although your PCR amplicons were already sent for sequencing, we will perform gel electrophoresis on the remaining sample to determine the quality of the PCRs.

1. Each group of four (4) will prepare a 1% agarose gel.
2. In a flask, combine 50 ml TBE buffer with 0.5 g agarose. Gently mix the solution.
3. Microwave the buffer-agarose mixture at 30 sec intervals, mixing the solution between heating. **Make sure you wear protective gloves as the flask will be hot!** The agarose should be completely dissolved in the buffer and be clear.
4. Add 10 ul SYBR Safe DNA stain to the hot flask and gently mix.
5. Place a comb into each gel apparatus. **Make sure the comb with 10 wells is placed facing down.** Let the hot gel mixture cool a bit before pouring (~3 min). Carefully pour the gel into the gel tray, making sure that no liquid is spilled.
6. Gels may be placed in the refrigerator to expedite solidification. This should take no more than 20 min.
7. When adequately solidified, place gel in gel apparatus. Make sure that the wells of the gel are near the negative (black) electrode. Fill gel apparatus with 1X TBE buffer. Chambers should be filled so that buffer adequately covers the gel.
8. Load the remainder of each sample into the gel. In addition, load 5 ul of the 100 bp ladder into the first well of each gel. This will help determine if the correct size fragment was obtained (~700 bp).

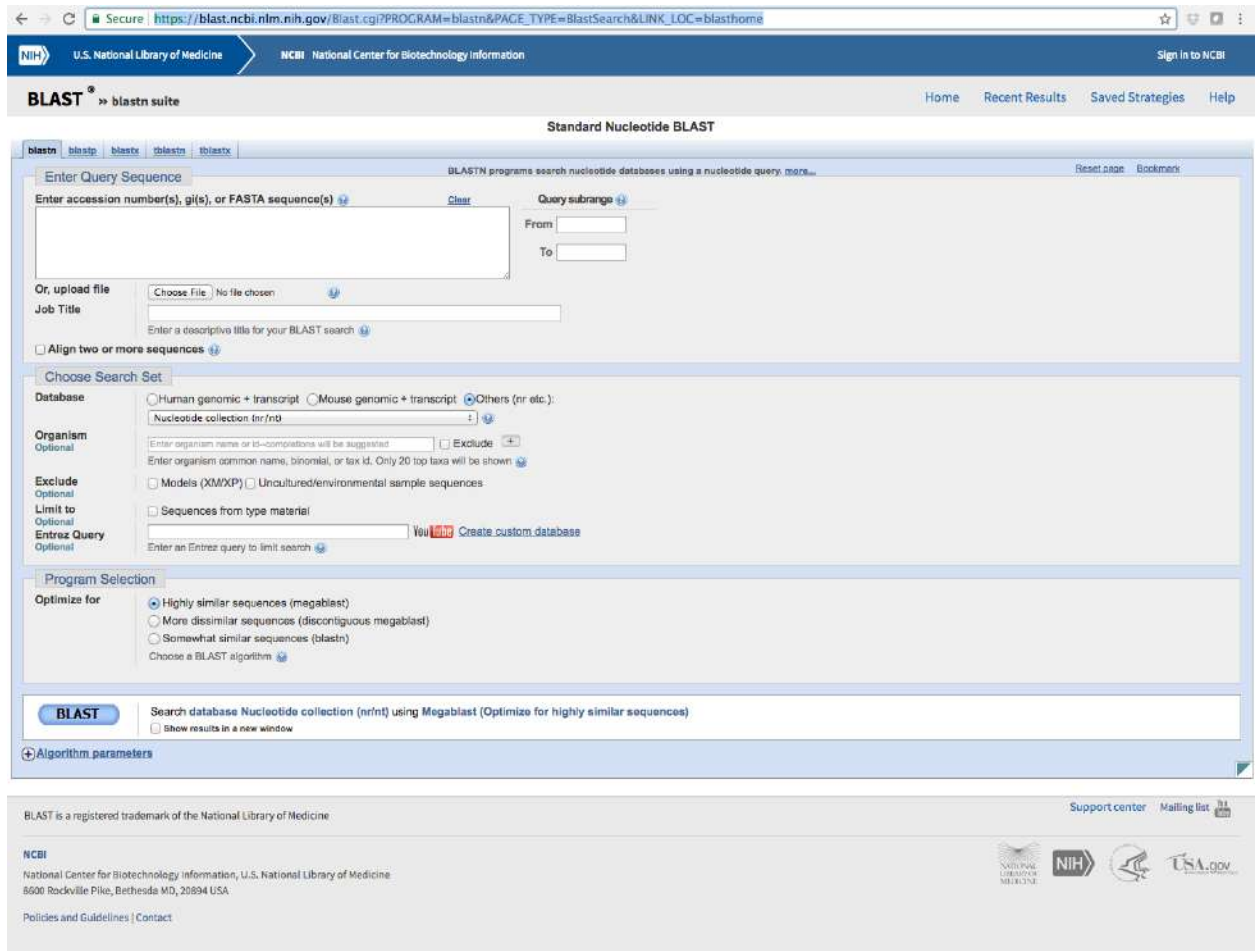
Sequence analysis

1. Your instructor will provide you with the DNA sequence data for the entire class. As we did with the PTC data, the first step will be to use FinchTV to edit the raw sequences to remove low quality bases. Again, these occur mostly at the beginning and end of the reads. However, it is important to manually scan the entire sequence to make sure that no base call is ambiguous. As this is a mtDNA gene there should



be no heterozygous peaks present. Once your sequences are edited, save them as fasta files (use the Export option).

2. Now that the data are cleaned, we are ready to determine the identity of our fish species. We will use NCBI's BLAST (Basic Local Alignment Search Tool) algorithm for this task.
3. Navigate to the following website (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) and paste each sequence into the box near the top (see image below).



4. Make sure that 'nucleotide collection' is selected under Database and click the BLAST button.
5. Scroll down until you see something resembling the image below. The 'Query cover' represents the proportion of your query sequence covered by the sequence match. For example, 100% of your query sequence aligns with the five samples in the figure below. The 'Ident' column indicates the sequence similarity between your query sequence and the matches. The five sequences below are 100% identical to your query sequence, suggesting that the species is *Salmo trutta*. Finally, the 'E value'



column represents the number of matches expected simply by random chance (i.e. **NOT** due to homology/shared ancestry). Lower values indicate true matches due to homology.

Descriptions

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Salmo trutta voucher NEFC_F16-334 mitochondrion, complete genome	1210	1210	100%	0.0	100%	MF821763.1
<input type="checkbox"/> Salmo trutta voucher NEFC_F16-333 mitochondrion, complete genome	1210	1210	100%	0.0	100%	MF821762.1
<input type="checkbox"/> Salmo trutta voucher NEFC_F16-157 mitochondrion, complete genome	1210	1210	100%	0.0	100%	MF821761.1
<input type="checkbox"/> Salmo trutta strain NEFC_F16-147 mitochondrion, complete genome	1210	1210	100%	0.0	100%	MF821760.1
<input type="checkbox"/> Salmo trutta voucher NXG2013304 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondriai	1210	1210	100%	0.0	100%	KX145518.1

- As we will be performing an evolutionary analysis of our fish sequences, we will need to obtain an **outgroup sequence to root our phylogenetic tree**. Outgroups are used to provide direction of evolutionary changes from ancestors to descendants. Outgroups should be from taxa closely related to the **ingroup**, but outside of the ingroup. Because we are working with fish species, a good outgroup might be from a mammal, bird, snake, or lizard. However, some students bring in shark tissue for analysis. In this case, using a tetrapod (amphibian, reptile, bird, or mammal) would not be appropriate because bony fish are more closely related to tetrapods than they are to cartilaginous fishes (sharks, skates, rays). Thus, for simplicity, we will use a lamprey as our outgroup. In NCBI, search for accession number **HQ686301.1**, which is a COI sequence from the Far Eastern Brook Lamprey, *Lethenteron reissneri*. Click on the accession to bring up a page like the one shown below. This page provides a ton of information about the sequence, such as the authors, publication, translation, etc. When ready, click the 'Send to' icon on the top right, send to file, and choose FASTA format. This will download the sequence in the same format as our fish sequences.



Nucleotide Advanced Help

COVID-19 Information Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

GenBank - Send to: - Change region shown Customize view

Lethenteron reissneri isolate Kalkaveim-21 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial

GenBank: HQ686301.1
[FASTA](#) [Graphics](#) [PopSet](#)

Go to:

LOCUS HQ686301 1072 bp DNA linear VRT 25-JUL-2016
 DEFINITION Lethenteron reissneri isolate Kalkaveim-21 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial.
 ACCESSION HQ686301
 VERSION HQ686301.1
 KEYWORDS .
 SOURCE mitochondrion Lethenteron reissneri (Far Eastern brook lamprey)
 ORGANISM [Lethenteron reissneri](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Cyclostomata; Hyperoartia; Petromyzontiformes; Petromyzontidae; Lethenteron.
 REFERENCE 1 (bases 1 to 1072)
 AUTHORS Artamonova,V.S., Kucheryavyy,A.V. and Pavlov,D.S.
 TITLE Nucleotide sequences of the mitochondrial cytochrome oxidase subunit I (COI) gene of lamprey classified with Lethenteron camtschaticum and the Lethenteron reissneri complex show no species-level differences
 JOURNAL Dokl. Biol. Sci. 437 (1), 113-118 (2011)
 PUBMED [21562960](#)
 REFERENCE 2 (bases 1 to 1072)
 AUTHORS Artamonova,V.S., Kucheryavyy,A.V. and Pavlov,D.S.
 TITLE Direct Submission
 JOURNAL Submitted (06-DEC-2010) Russian Academy of Science, A.N. Severtsov Institute of Ecology and Evolution, 33 Leninskij Prosp., Moscow 119071, Russia

FEATURES
 source Location/Qualifiers
 1..1072
 /organism="Lethenteron reissneri"
 /organelle="mitochondrion"
 /mol_type="genomic DNA"
 /isolate="Kalkaveim-21"
 /isolation_source="resident lamprey"
 /db_xref="taxon:7753"
 /country="Russia: Kamchatka peninsula, r. Kalkaveyem (tributary of r. Utkholok)"
 /collection_date="2006"
 /collected_by="A.V. Kucheryavyy"
 gene
 <1..>1072
 /gene="COI"
 CDS
 <1..>1072
 /gene="COI"
 /codon_start=2
 /transl_table=2
 /product="cytochrome oxidase subunit I"
 /protein_id="AEG75867.1"
 /translation="MMLSPDMAFPRMNMMSFWLLPPSLLLLASAGVEAGAGTGWTVYPLAGNLAHTGASVDLTI FSLHLAGISSILGAVNFITTFNMKPTMTQYQTPLEW SVLITAVLLLSLPVLA AAITMLTDRNLNTSFFDPAGGGDPILYQHLFWFFGHPEVY ILILPGFGIISHVVAYYS GKKEPFYMGVMWAMMAIGLLGFIVWAHMFTVGHMDVDR AYFTSATMIIAIPGVKVFVSWLATLHGKIVWHTPMLWALGFIFLFTVGGTLGIVLSN SSDLIILHDITYYVAHFHYVLSMGAVFAIMAGFVHWPLFTGYTLNETWSKAHFVIMF TGVNLTFFPOHFLGLAGMPRRYS"

ORIGIN
 1 tataaactt agcggccctg atatagcctt cccacgata aataacataa gcttttgact
 61 gctcccacca tcctactct tacttttagc ttccgcagga gttgaagcag gaggccggaac
 121 tggatgaaca gtataccac cctcagcagg aaatttagcc cacacagggg cctctgttga
 181 cttaaacatt ttctcccttc atctagccgg tatttcatca atccttgggg cagtcaactt
 241 tattacaaca attttaaaca taaaacctcc aactataaca caatacacia cccattatt
 301 tgtatgatcc gttttaatta ctgcagtcct cctcttctca tcacttctgt tacttgagc
 361 tgccatcact atacttttaa cagatcgtaa ttaataaca tccttctttg acctgcagg
 421 aggaggagac ccaatccttt accaacacct gttctgattc ttggggcacc ctgaagtta
 481 tattctaatt ttaccaggct ttggaattat ctctcagta gtcgctact actccggga
 541 aaaagaacca ttggatata taggaatagt ctgagcaata atagccattg ggttactagg
 601 gttatttga tgagccacc acatattcac ggttaggaata gatgtgaca cagagccta
 661 ctttacatca gccacaataa ttattgctat tccaacagga gttaaagtct tttagctgact
 721 agccacactc catggcggaa aatcgtgtg acacacccc atactatgag ccctaggctt
 781 cattttctta ttactgtag ggggacttac cggaattgtc ttatccaact catcactaga
 841 tattacctt catgatactt attatgtagt agctcacttc cattatgata tctctatagg
 901 agctgttttc gcaatcatag ctggcttctg tcattgattc ccactatta caggatatac
 961 acttaacgaa acctgatcaa aagcacactt tgtcattata ttactgggtg taaacttacc
 1021 attcttccct caacatttcc taggttttagc tggatataca cgacttact ca

//

Related information
 Protein
 PubMed
 Taxonomy
 PopSet

Recent activity Turn Off Clear
 Lethenteron reissneri isolate Kalkaveim-21 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial
 lamprey coi (373)

[See more...](#)



7. Next, import all the fish data and the outgroup sequence into Alivew to perform a **multiple sequence alignment**. Multiple sequence alignments are needed prior to evolutionary inference to determine homology among sequences. In essence, we need to make sure we are comparing apples to apples among our different species. Once all the data are imported into Alivew, click Align > realign everything. This will use the popular MUSCLE algorithm to align all your sequences. Once the data are aligned, export the alignment as a **phylip** file.
8. Our final objective is to perform a phylogenetic analysis of our sequences to determine evolutionary relationships. There are multiple optimality criteria used to build phylogenetic trees including **maximum parsimony (MP)**, **maximum likelihood (ML)**, and **Bayesian inference (BI)**, the latter two of which are considered parametric methods because they rely on evolutionary models. Before we estimate a tree, we need to become a little familiar with what evolutionary models are.

Brief introduction to evolutionary models:

Evolutionary models are used to model changes in characters states (e.g. nucleotides) over the course of evolutionary time. For example, what is the probability of an adenine changing to a thymine versus a cytosine? These models can be used to estimate phylogenetic trees using maximum likelihood and Bayesian methods. So what exactly IS an evolutionary model? All evolutionary models have two components: base frequencies and nucleotide substitution rates. The simplest model, the Jukes-Cantor (JC) model, assumes that the four nucleotides occur at the same frequency in the alignment (25% each) and that the rate of nucleotide substitution is the same between any pair of nucleotides (e.g. $A > C = A > T = G > T = C > T$ and so forth). The Felsenstein 81 (F81) model is the same as the JC model except it assumes unequal base frequencies ($\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$). The Kimura two parameter model (K2P) assumes equal base frequencies ($\pi_A = \pi_C = \pi_G = \pi_T$), but differentiates between the rate of transitions and transversions. Recall that a transition is between a purine to purine or between a pyrimidine to pyrimidine. In contrast, transversions are between a purine and pyrimidine. The HKY85 model expands the K2P model to include unequal base frequencies. The most parameter rich model is the General Time Reversible (GTR) model, which assumes unequal base frequencies and different substitution rates between every nucleotide pair. In general, more parameter rich models like HKY and GTR are needed for alignments with higher levels of polymorphism.

In addition to the classes of models discussed above, we can add two additional parameters to any model. The first accounts for invariable sites in a multiple sequence alignment, which we denote as I . The second parameter, gamma, controls for evolutionary rate variation across the alignment. For example, some regions of the alignment might be highly conserved, whereas other portions show many nucleotide substitutions. The amount of rate variation is controlled by a parameter alpha, with small values indicating high levels of rate variation across the alignment. These additional parameters (+I, +G) can be added to any of the models discussed above (e.g. HKY+I, GTR+I+G, JC+G, etc.).

It should be obvious by now that there are many potential models to be used to model the evolutionary process. So how do researchers know which model to use for a specific alignment? Fortunately, there are software packages available

that can help select a model objectively. We will not go into model selection in further detail here as it is beyond the scope of the course.

Now that we are a little more comfortable with evolutionary models, let's get back to our fish data. We just exported our multiple sequence alignment from Aliview as a phylip file. We will use the popular program **IQ-TREE** to estimate a phylogeny using the **ML criterion**. Navigate to www.iqtree.org to download the program to your computer. There are versions available for both Windows and Mac. IQ-TREE, and most phylogenetics packages, are command-line only and thus require some familiarity with the terminal. Fortunately, in IQ-TREE the commands are nearly identical between operating systems. One difference is that in Windows we use '\ ' in our directory paths, whereas on a Mac we use '/ '.

Once you have the program downloaded, place your phylip alignment file in the same directory as IQ-TREE. Next, we need to open up a terminal window. For Windows, click 'Start' and type "cmd". This will open up a command line interface for your analysis. On a Mac, click the little magnifying glass on the upper right of your screen and type "terminal".

Next, we need to **change directory** to where the program and data are. In both Windows and Mac we do this using the **cd** command. For example,

```
cd Documents/Programs/iqtree (for Mac)
```

```
cd Documents\Programs\iqtree (for Windows)
```

We are now ready to run a ML phylogenetic analysis! IQ-TREE has many great options, but we will keep it basic here. Assuming you are on a Mac and in the iqtree directory, type the following command at the prompt:

```
./iqtree -s NameOfYourAlignment -B 10000 -alrt 1000 -o NameOfOutgroup
```

This command will perform multiple steps. First, IQ-TREE will search for the best nucleotide substitution model using ModelFinder. Next, it will use the best model to infer the ML phylogeny. Finally, it will assess node support using both the ultrafast bootstrap and SH-like approximate likelihood ratio tests. In general, ultrafast bootstrap values >95% and SH-aLRT > 80% indicate strong support for relationships. When the analysis finishes, you will see multiple output files in your working directory. Your instructor should take some time to go over these with you. The file ending with **.treefile** is your ML phylogeny.

Our next step is to view our fish phylogeny. Unfortunately, IQ-TREE is not the best place to view the tree. Instead, we will use the program **FigTree** (<https://github.com/rambaut/figtree/releases>). Download the relevant version for your system and open the program. Next, click 'File/Open' and navigate to your **.treefile** file. Click OK and your phylogeny should pop up. The first thing to do is root your phylogeny with the outgroup. Click on the branch leading to the lamprey and then click the 'Reroot' button in the menu. The next thing we want to do is show the support values for the nodes. Click on the box for 'Branch Labels' and then choose 'label'. This will map on the



aLRT/ultrafast bootstrap values on branches. Remember that values greater than 80% (aLRT) / 95% (bootstrap) indicate strong support. There are many other options worth exploring in FigTree. When ready, export your phylogeny as a pdf file.

Review questions

1. Paste your fish phylogeny in the space below and describe the phylogeny in your own words. Which species are sister taxa? Which are distantly related? Do your inferred relationships appear to agree with current taxonomy? If your samples included multiple species within a single genus (e.g. *Salmo*), was the genus monophyletic?



2. Discuss the utility of DNA barcoding for biodiversity assessment and conservation. How do you think the approach has advanced conservation efforts? How is DNA barcoding better than traditional morphological identification? What might be some limitations to barcoding?

3. As stated above, the COI gene is usually not an ideal marker for barcoding plants. Why might this be? What is different between the plant barcoding genes and COI?

4. In BLAST, what does the E-value indicate?



9. Discuss how the statistical procedure of bootstrapping is used in phylogenetic inference. What is meant by a bootstrap replicate?

References

Arnold ML, Holman D, Zweifel S. 2017. Using molecular biology and bioinformatics to investigate the prevalence of mislabeled fish samples. *The American Biology Teacher* 79, 763-768.

Warner K, Timme W, Lowell B, Hirschfield M. 2013. Oceana study reveals seafood fraud nationwide. Retrieved from http://oceana.org/sites/default/files/reports/National_Seafood_Fraud_Testing_Results_Final.pdf.

